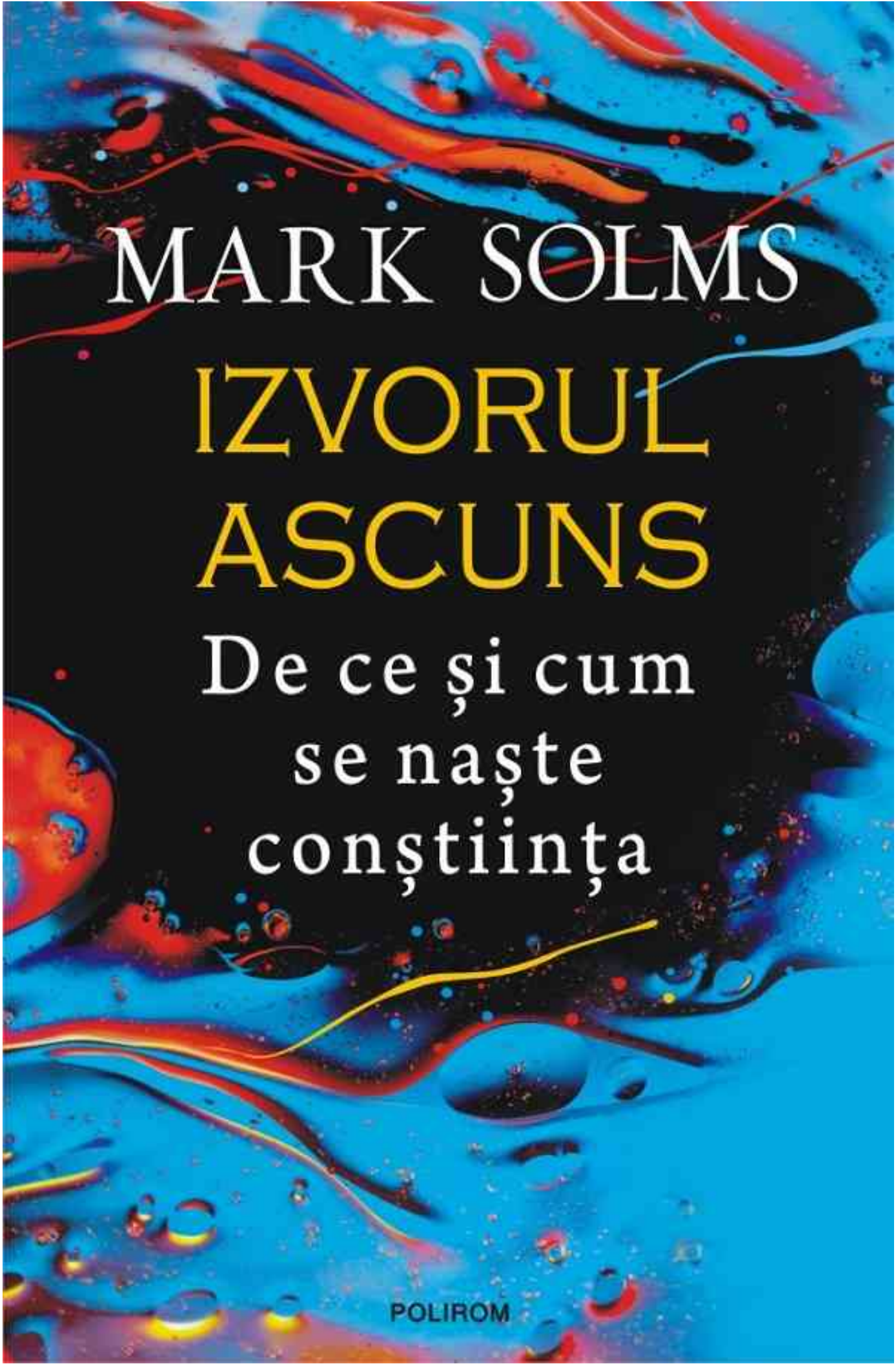

MARK SOLMS
IZVORUL ASCUNS
De ce și cum se naște conștiința
POLIROM

MARK SOLMS

IZVORUL ASCUNS

De ce și cum se naște conștiința

Traducere de Dan Bălănescu

POLIROM
2022

MARK SOLMS, psihanalist și neuropsiholog, a descoperit mecanismele cerebrale ale viselor. Este cercetător în cadrul departamentului de neuropsihologie de la Neuroscience Institute, University of Cape Town. În anul 2000 a înfiinţat International Neuropsychoanalysis Society. Este conferenţiar onorific în neurochirurgie al Royal London Hospital School și membru de onoare al American College of Psychiatrists. A mai publicat *The Brain and the Inner World* (împreună cu Oliver Turnbull) (2002).

În amintirea lui
Jaak Panksepp
(1943-2017)

A rezolvat vechea ghicitoare și a fost un om foarte înțelept.

Lista figurilor

1. Principalele structuri cerebrale implicate în somnul REM.
2. Principalele structuri cerebrale implicate în vis
3. Scanări PET ale creierului la debutul somnului și în timpul somnului REM
4. RMN al unui copil născut fără cortex cerebral (Copyright © ACR Learning File, Neuroradiology, Edition 2, 2004, © American College of Radiology)
5. Reacție emoțională a unui copil născut fără cortex cerebral (figură din „Consciousness without a cerebral cortex: A challenge for neuroscience and medicine” de Bjorn Merker, *Behavioral and Brain Sciences*, 1 mai 2007, Cambridge University Press © 2007 Cambridge University Press)
6. Coliculii superiori, cortexul vizual și PAG
7. RMN-uri ale unui bărbat prezentând distrugere completă a lobilor prefrontali
8. Prima diagramă a sistemelor memoriei realizată de Freud
9. Imagistică RMN a creierului în patru stări emoționale
10. Tipare tipice ale undelor activității corticale înregistrate cu ajutorul EEG
11. Reprezentare grafică a măsuritorilor potențialelor de acțiune (*spike trains*) a 20 de neuroni (figură din „Simulating Neural Spike Trains” de Praneeth Namburi,

Praneethnamburi.com, 5 februarie 2015 © Praneeth Namburi 2015)

12. Homeostaza sentimentului
13. Sistem cu autoorganizare și zona sa Markov (figură din „Life as We Know It” de Karl Friston, *Journal of the Royal Society Interface*, 10, 6 septembrie 2013, https://royalsocietypublishing.org/doi/10.1098/rsif.2013.0475, CC BY)
14. Efecte entropice ale unor leziuni la nivelul zonei Markov
15. Ierarhie predictivă simplificată
16. Reprezentare grafică a unei predicții posterioare (figură din Hidesyoshi Yanagisawa, Oto Kawamata și Kazutaka Ueda, „Modeling Emotions Associated with Novelty at Variable Uncertainty Levels: A Bayesian Approach”, *Frontiers in Computational Neuroscience*, 24 ianuarie 2012, https://www.frontiersin.org/articles/10.3389/fncom.2019.00002/full)
17. Dinamica unui sistem care se autovalidează, cu optimizarea preciziei
18. Testul figurilor complexe Rey aplicat unui copil cu vedere inversată
19. Schemă de învățare prin eroare de predicție a recompensei

Introducere

Când eram mic, mi-am pus o întrebare ciudată: cum ne putem imagina lumea care exista înainte de evoluția conștiinței? Desigur, a existat o astfel de lume, dar cum ți-o *imaginezi* – lumea așa cum era înainte ca reprezentarea lucrurilor să devină posibilă?

Ca să înțelegi ce vreau să spun, încearcă să-ți imaginezi o lume în care *un răsărit de soare nu poate avea loc*. Pământul s-a rotit întotdeauna în jurul Soarelui, dar soarele răsare la orizont doar din punctul de vedere al unui observator. Este un eveniment care ține inerent de perspectivă. Răsăritul soarelui va fi întotdeauna prins în capcana experienței subiective.

Această asumare obligatorie a perspectivei este ceea ce face atât de dificilă înțelegerea conștiinței. Dacă vrem să reușim, trebuie să eludăm subiectivitatea – să privim conștiința din exterior, să vedem lucrurile așa cum sunt cu adevărat, și nu cum ni se par. Dar cum facem asta? Cum ne desprindem de propriul sine?

În tinerețe, îmi vizualizam cu naivitate conștiința ca pe o bulă care mă înconjoară: conținutul ei era format din imaginile în mișcare, sunetele și alte fenomene ale experienței. Dincolo de această bulă, presupuneam că era un întuneric infinit. Îmi imaginam acest întuneric ca pe o simfonie a mărimilor

adimensionale, a forțelor interdependente, a energiilor și a altora asemenea: adevărata realitate „din exterior” pe care conștiința mea o reprezintă în formele calitative inevitabile.

Imposibilitatea oricărei astfel de imaginări – a reprezentării realității fără reprezentări – ilustrează amploarea sarcinii pe care mi-am propus-o în această carte. Încă o dată, după atâția ani, încerc să arunc o privire în spatele vălului conștiinței, să-i surprind mecanismele.

Din acest motiv, cartea de față e inevitabil perspectivală. De fapt, e chiar mai perspectivală decât i-o impune paradoxul pe care tocmai l-am descris. Ca să te ajut să vezi lucrurile din punctul meu de vedere, m-am hotărât să-ți împărtășesc o parte din povestea mea. Progresele în ideile mele științifice despre conștiință au apărut adesea în urma unor evenimente din viața mea personală și din activitatea mea clinică și, deși cred că concluziile mele sunt de sine stătătoare, e mult mai ușor să le înțelegi dacă știi cum am ajuns la ele. Unele dintre descoperirile mele – de exemplu, mecanismele cerebrale ale visului – au fost în mare parte întâmplătoare. Unele dintre alegerile mele din punct de vedere profesional – de exemplu, să mă abat puțin de la cariera mea în neuroștiințe și să mă formez ca psihanalist – mi-au adus mai multe satisfacții decât aș fi putut spera. În ambele cazuri, voi explica și cum.

Dar, în măsura în care misiunea mea de a înțelege conștiința şi-a atins scopul, adevăratul meu noroc este că am avut nişte colaboratori geniali. În special, am avut marele noroc să lucrez cu regretatul Jaak Panksepp, un specialist în neuroștiințe care, mai mult decât oricare altul, a înțeles originea și puterea *sentimentelor*. Cam tot ce cred eu astăzi despre creier a fost influenţat de ideile lui.

Mai recent, am avut posibilitatea să lucrez cu Karl Friston, care, printre numeroasele sale calități formidabile, are și renumele de a fi cel mai influent specialist în neuroștiințe în viață din lume. Friston a pus bazele cele mai solide ale teoriei pe care mă pregătesc să o elaborez. E celebru în principal pentru că a redus funcțiile cerebrale (de toate tipurile) la o necesitate fizică fundamentală de a minimiza ceva numit *energie liberă*. Explic acest concept în capitolul 7, dar, deocamdată, voi preciza doar că teoria la care am lucrat împreună cu Friston se leagă de acest proiect – în asemenea măsură încât poate fi numită teoria energiei libere a conștiinței. Despre asta e vorba.

Explicația finală a capacității de a simți este o enigmă atât de complicată, încât astăzi este numită cu respect „problema dificilă”. Uneori, după rezolvarea unei enigme, atât întrebarea, cât și răspunsul încetează să mai fie interesante. Rămâne la latitudinea ta să stabilești dacă ideile pe care le prezint aici aruncă o lumină nouă asupra acestei probleme dificile. În orice caz, sunt convins că te vor ajuta să te vezi pe *tine însuți* într-o lumină nouă și, în acest sens, ar trebui să rămână interesante până când sunt schimbate. Până la urmă, într-un sens profund, tu *ești* conștiința ta. Așadar, pare rațional să te aștepți ca o teorie a conștiinței să explice fundamentele motivelor pentru care simți așa cum simți. Ar trebui să explice de ce ești așa cum ești. Poate că ar trebui să clarifice chiar și ce poți face în această privință.

Recunosc că ultima temă trece dincolo de ce mi-am propus legat de sfera acestei cărți. Dar nu și dincolo de sfera teoriei. Ceea ce spun eu despre conștiință reunește într-o singură poveste fizica elementară a vieții, cele mai recente progrese în

neuroștiința computațională și în cea afectivă, precum și subtilitățile experienței subiective analizate de obicei de psihanaliști. Cu alte cuvinte, lămuririle aduse de această teorie ar trebui să-ți fie folositoare.

Această teorie e munca mea de-o viață. După multe decenii, încă mă întreb cum ar fi putut arăta lumea înainte să fi existat cineva care să o vadă. Acum, că dețin mai multă cunoaștere, îmi imaginez zorii vieții într-un izvor geotermal. Organismele unicelulare ce au luat ființă acolo cu siguranță nu au fost conștiente, dar șansele lor de supraviețuire au fost afectate de mediul ambiant. E ușor să ne imaginăm aceste organisme simple răspunzând „bunăvoinței" biologice a energiei solare. De acolo, nu mai e decât un pas până la a ne imagina creaturi mai complexe care caută activ asemenea surse de energie și care, în cele din urmă, dezvoltă o capacitate de a evalua șansele de succes prin acțiuni alternative.

În opinia mea, conștiința a luat naștere din experiența unor asemenea organisme. Imaginează-ți arșița zilei și frigul nopții din perspectiva acestor prime ființe. Valorile fiziologice care le înregistrau experiențele diurne au fost precursoarele primului răsărit de soare.

Mulți filosofi și oameni de știință încă cred că simțirea nu servește niciunui scop fizic. Sarcina mea în această carte este să te conving că o interpretare alternativă este plauzibilă. Asta mă obligă să te conving că sunt parte a naturii, că nu sunt fundamental diferite de alte fenomene naturale și că ele *fac* ceva în interiorul matricii cauzale a lucrurilor. Voi demonstra că conștiința este legată de faptul de a simți, iar faptul de a simți, la rândul său, are legătură cu cât de bine sau de prost te descurci în viață. Conștiința există ca să te ajute să te descurci

mai bine.

Se spune că problema dificilă a conștiinței este cea mai mare enigmă nerezolvată a neuroștiințelor contemporane, dacă nu chiar a științei în general. Soluția propusă în această carte se îndepărtează radical de abordările convenționale. Întrucât cortexul cerebral este sediul inteligenței, aproape toate lumea crede că e și sediul conștiinței. Eu nu sunt de acord – conștiința este mult mai primitivă. Ea apare într-o zonă a creierului pe care o au şi oamenii, și peștii. Acesta este „izvorul ascuns” din titlu.

Conștiința nu trebuie confundată cu inteligența. Poți să simți durere fără să ai vreo idee despre cauza ei. De asemenea, nu e neapărat ca impulsul de a mânca – senzația de foame – să implice vreo comprehensiune intelectuală a exigențelor vieții. Conștiința în forma ei elementară, şi anume *sentimentele* brute, este o funcție surprinzător de simplă.

Regăsim această abordare şi la alți trei specialiști de marcă în neuroștiințe: Jaak Panksepp, Antonio Damasio și Björn Merker. Panksepp a fost deschizător de drumuri. El (ca şi Merker) a făcut cercetări pe animale, Damasio (ca şi mine) nu. Mulți cititori se vor îngrozi de concluziile studiilor pe animale pe care le voi prezenta aici, tocmai deoarece demonstrează că alte animale simt la fel ca noi. Toate mamiferele simt durere, frică, panică, tristețe și altele asemenea. E o ironie că tocmai cercetările lui Panksepp au eliminat orice urmă de îndoială în această privință. Singura noastră consolare este că descoperirile lui au făcut imposibilă continuarea în aceeași manieră a unor asemenea studii.

Ideile lui Panksepp, Damasio și Merker m-au atras deoarece

aceștia cred, ca și mine, că neuroștiințelor actuale le lipsește o focalizare clară pe natura concretă a *experienței trăite*. Se poate spune că ceea ce ne unește este că am clădit, uneori fără să ne dăm seama, pe temeliile abandonate pe care le-a turnat Freud pentru o știință a gândirii care pune simţămintele înaintea cogniției. (Cogniția este în cea mai mare parte inconștientă.) Acesta e al doilea sens în care volumul de față se desprinde radical de teoriile general acceptate – se întoarce la „Proiectul" lui Freud din 1895 și încearcă să-i duc sarcina la bun sfârșit. Dar nu îi trec cu vederea numeroasele erori. De exemplu, ca toată lumea, Freud credea că conștiința este o funcție corticală.

În al treilea rând, volumul de față ajunge la concluzia că conștiința se pretează la inginerii. Că poate fi *produsă* artificial. Această concluzie, cu profundele sale implicații metafizice, izvorăște din colaborarea mea cu Karl Friston. Spre deosebire de Panksepp, Damasio și Merker, Friston este specialist în neuroștiințele computaționale. Așadar, el crede că conștiința este în esenţă reductibilă la legile fizicii (convingere în mod surprinzător împărtășită de Freud). Însă chiar și Friston a pus în mare parte semnul egalității între funcțiile psihice și cele corticale înainte să ne începem colaborarea. Această carte duce cadrul statistico-mecanic al lui Friston la un nivel mai profund, în cele mai primitive cotloane ale trunchiului cerebral…

Aceste trei perspective diferite fac problema dificilă mai puțin dificilă. Iar cum anume va explica volumul de față.

Mark Solms
Chailey, East Sussex
martie 2020

1

Materia viselor

M-am născut pe Coasta Scheletelor din fosta colonie germană Namibia, unde tatăl meu administra o mică firmă sud-africană numită Consolidated Diamond Mines. Compania-mamă, De Beers, crease practic un stat în stat, numit *Sperrgebiet* („zonă interzisă”). Minele sale aluvionare se întindeau de la dunele de nisip ale Deșertului Namib până pe fundul Oceanului Atlantic, pe o distanță de mai mulți kilometri în larg.

Acesta a fost peisajul aparte care mi-a modelat imaginația. În copilărie, mă jucam cu fratele meu mai mare Lee de-a minele de diamante, folosind excavatoare de jucărie, recreând în grădina noastră impresionantele realizări inginerești la care eram martori când ne ducea tata să vedem exploataţiile miniere de suprafaţă din deșert. (Eram, desigur, prea mici ca să cunoaștem aspectele mai puțin impresionante ale ocupației lui tata.)

Într-o zi din 1965, când aveam patru ani, părinții mei se plimbau cu iahtul la Cormorant Yacht Club, cum făceau adesea, iar eu mă jucam în club cu Lee, care avea şase ani. Soarele arzător risipise ceața de la primele ore ale dimineții. Am ieșit

din interiorul răcoros al clădirii de trei etaje a clubului, îndreptându-mă spre malul apei. Mă plimbam cu picioarele în apă și urmăream niște peștișori sclipitori cum se îndepărtează de degetele mele, în timp ce Lee și niște prieteni de-ai lui s-au cățărat pe acoperiș.

Am trei flashbackuri. Primul, cu sunet ca pârâitul unui pepene verde care se crapă. Al doilea, imaginea lui Lee întins la pământ și smiorcăindu-se că îl doare un picior. Ultimul, mătușa și unchiul spunându-mi că vor avea ei grijă de mine și de sora mea cât părinții noștri merg la spital cu Lee. Partea cu piciorul dureros e probabil o închipuire: în dosarul medical scrie că fratele meu și-a pierdut cunoștința la impactul cu trotuarul de beton.

Lee avea nevoie de un tip de îngrijiri de specialitate pe care spitalul nostru local nu le putea oferi. A fost dus cu elicopterul la spitalul Groote Schuur din Cape Town, la 800 de kilometri distanță. Pe vremea aceea, secția de neurochirurgie funcționa într-o clădire impresionantă construită în stil Cape Dutch, exact aceeași clădire în care lucrez eu acum ca neuropsiholog. Lee își fracturase craniul și suferise o hemoragie intracraniană. Când se extind astfel de hematoame, ele constituie o urgență ce amenință viața și necesită intervenție chirurgicală. Fratele meu a avut noroc: și-a revenit în următoarele câteva zile și, în cele din urmă, a fost externat și trimis acasă.

În afară de faptul că după accident a trebuit să poarte o cască de protecție pentru craniul fracturat, Lee nu arăta diferit. Totuși, persoanalitatea lui se modificase profund. Germana are un cuvânt pentru senzația pe care mi-a provocat-o acest lucru, *Unheimlichkeit*, pentru care nu există un echivalent exact în engleză. Literal, înseamnă „ciudăţenie”, dar s-ar traduce mai

bine prin „ceva care îți dă fiori reci” sau „stranietate”.

Cel mai evident, s-a schimbat în sensul că și-a pierdut achizițiile ce marchează trecerea într-o nouă etapă de dezvoltare. O vreme, și-a pierdut parțial chiar şi controlul sfincterelor. Ceea ce m-a tulburat mai mult a fost faptul că părea să *gândească* altfel decât înainte. Parcă Lee era și în același timp nu era prezent. Parcă uitase multe dintre jocurile pe care le jucam. Acum, jocul nostru de-a săpatul după diamante a devenit pur și simplu săpat gropi. Aspectele imaginative și simbolice ale jocului nu îl mai atrăgeau. Nu mai era Lee.

Nu a trecut clasa – era primul său an de școală. Din acele prime zile după accident îmi amintesc mai des încercarea mea de a mă împăca cu dihotomia că fratele meu care se întorsese acasă arăta la fel, dar nu era la fel. Mă întrebam unde dispăruse versiunea sa anterioară.

În anii ce au urmat am căzut în depresie. Îmi amintesc că nu reușeam să mă mobilizez ca să mă încalț dimineața, să merg la școală. Asta se întâmpla la vreo trei ani după accident. Nu îmi găseam energia să fac aceste lucruri pentru că nu le mai vedeam rostul. Dacă însăși ființa noastră depinde de funcționarea creierului, oare ce mi se va întâmpla când va muri creierul *meu*, ce se va întâmpla cu restul trupului? Dacă mintea lui Lee era cumva reductibilă la un organ din corp, cu siguranță că și a mea era. Asta însemna că eu – ființa mea conştientă – voi exista doar o perioadă relativ scurtă. Apoi voi dispărea.

Întreaga mea carieră științifică m-am tot gândit la această problemă. Am vrut să înțeleg ce i s-a întâmplat fratelui meu și ce ni se va întâmpla în timp tuturor. Aveam nevoie să înțeleg, în termeni biologici, ce înseamnă existența noastră ca ființe care

trăiesc ceva. Pe scurt, să înțeleg conștiința. Din acest motiv m-am specializat în domeniul neuroștiințelor.

Chiar și astăzi cred că nu aș fi putut urma un drum mai direct spre răspunsurile pe care le căutam.

S-ar putea ca natura conștiinței să fie cel mai dificil subiect al științei. El contează pentru că fiecare *este* conștiința lui, dar e controversat din cauza a două enigme ce îi chinuie pe gânditori de secole. Prima este întrebarea care e legătura dintre minte și corp – sau, pentru cei cu o înclinație materialistă (adică aproape toți specialiștii în neuroștiințe), cum creează creierul mintea. Aceasta este numită „problema minte/corp". Cum produce creierul fizic experiența fenomenală? Și încă o întrebare la fel de derutantă: cum reușește chestia aia nonfizică numită conștiință să controleze corpul fizic?

Filosofii încadrează această problemă în domeniul pe care ei îl numesc „metafizică", adică un fel de a spune că ei nu cred că poate fi rezolvată din punct de vedere științific. De ce nu? Deoarece știința se bazează pe metode empirice, iar „empiric" implică „derivat din dovezi senzoriale". Mintea nu este accesibilă observației senzoriale. Nu poate fi văzută sau atinsă, este invizibilă și intangibilă, este o temă, nu un obiect.

A doua enigmă se referă la ce putem ști despre minte din exteriorul ei – la drept vorbind, cum ne putem da seama că mintea există. Această enigmă este numită „problema altor minți", adică, dacă mințile sunt subiective, atunci nu o poți observa decât pe a ta. Și atunci cum putem să ne dăm seama dacă alți oameni (sau alte ființe sau mașini) au minte, darămite să identificăm legi obiective care guvernează modul în care funcționează în general mințile?

În ultimul secol, aceste întrebări au primit trei răspunsuri științifice importante. Știința se bazează pe experimente. Un lucru în avantajul nostru este că metoda experimentală nu aspiră la adevăruri supreme, ci mai degrabă la ceea ce poate fi descris drept cele mai bune presupuneri. Pornind de la observații, facem conjecturi în legătură cu explicațiile plauzibile ale fenomenelor observate. Cu alte cuvinte, formulăm ipoteze. Apoi generăm *predicții* pornind de la ipoteze. Acestea iau forma: „Dacă ipoteza X este corectă, atunci, când fac Z, ar trebui să se întâmple Y” (în care există o șansă rezonabilă ca Y să nu se întâmple dacă se pleacă de la alte ipoteze). Acesta este experimentul. Dacă Y nu se întâmplă, atunci se deduce că X este falsă și este revizuită în conformitate cu observațiile noi. Apoi, procesul experimental o ia de la capăt, până când dă naștere unor predicții falsificabile care sunt confirmate. În acest moment, considerăm că ipoteza este *provizoriu* adevărată, până când și dacă nu o contrazic alte observații. În acest fel, nu ne așteptăm să ajungem la certitudine în știință, ci aspirăm doar la mai puțină incertitudine[1].

Începând din prima jumătate a secolului XX, o școală de psihologie numită „behaviorism” a început să aplice minții în mod sistematic metoda experimentală. Punctul său de plecare era să elimine din câmpul cercetării tot, cu excepția evenimentelor observabile în mod empiric. Behavioriștii au eliminat toate discuțiile „mentaliste” despre credințe și idei, sentimente și dorințe și și-au limitat domeniul de studiu la răspunsurile vizibile și tangibile ale subiectului la stimuli obiectivi. Respingeau cu fanatism descrierile subiective a ceea ce se petrece în interior. Tratau mintea ca pe o „cutie neagră”,

despre care nu se puteau ști decât inputurile și outputurile.

De ce au adoptat o poziție atât de extremă? Bineînțeles, aceasta era parțial o încercare de a ocoli problema altor minți. Dacă refuzau de la bun început să accepte orice discuție despre alte minți, se înţelegea de la sine că teoriile lor nu puteau fi afectate de îndoielile filosofice atât de răspândite în psihologie. De fapt, au exclus psihicul din psihologie.

Acesta poate părea un preț cam mare. Dar behaviorismul a fost de la început o doctrină revoluționară. Behavioriștii nu urmăreau puritatea epistemologică ca scop în sine, ci încercau și să detroneze puterea dominantă în psihologie la vremea respectivă. Psihanaliștii freudieni dominaseră știința minții de la începutul secolului. Examinând îndeaproape caracteristicile neobișnuite ale mărturiilor introspective, Sigmund Freud căutase să dezvolte un model al minții analizate, cum ar veni, dinspre interior spre exterior. Ideile rezultate au stabilit agenda cu privire la tratament și cercetare pentru o jumătate de secol, generând instituții, experți acreditați și un nucleu de intelectuali deschizători de drumuri remarcabili. Totuși, behavioriștii considerau că toate teoriile lui Freud nu erau decât nişte fantezii brodate pe bazele fragile ale subiectivității. Freud se aruncase cu capul înainte în problema altor minți și târâse după el şi restul psihologiei. Behavioriștii și-au propus să o scoată la liman.

În pofida austerității programului lor, ei chiar au reușit să deducă relații cauzale dintre anumite tipuri de răspunsuri și stimuli mintali. Și nu numai: au reușit să manipuleze inputurile astfel încât să obțină schimbări previzibile la nivelul outputurilor. Astfel, au descoperit unele dintre legile fundamentale ale învățării. De pildă, când declanșatorul unui

comportament involuntar este însoțit în mod repetat de un stimul artificial, atunci stimulul artificial va ajunge să declanșeze același răspuns involuntar ca stimulul natural. Așadar, dacă imaginea hranei este însoțită în mod repetat de sunetul unui clopoțel (la animalele care salivează în mod natural atunci când văd hrană, precum câinii), atunci sunetul clopoțelului va ajunge să declanșeze salivarea şi singur. Aceasta se numește „condiționare clasică”. De asemenea, dacă un comportament voluntar este însoțit în mod repetat de recompense, comportamentul respectiv va apărea mai frecvent, iar dacă același comportament este însoțit de pedepse, se va manifesta mai rar. Deci, dacă un câine ce sare pe oaspeți primește îmbrățișări, va sări și mai mult pe ei; dacă primește o lovitură, va sări mai puțin. Acest lucru se numește „condiționare operantă” – sau legea efectului.

Asemenea descoperiri nu au fost realizări mărunte. Ele au demonstrat că mintea se supune unor legi naturale, ca orice altceva. Dar mintea înseamnă mult mai mult decât învățarea și chiar învățarea este influențată și de alți factori în afară de stimulii externi. Imaginează-ți că îți spui: „După ce citesc această pagină, îmi fac un ceai”. Acest tip de gândire îți influențează comportamentul tot timpul. Behavioriștii nu considerau însă astfel de rapoarte introspective date științifice acceptabile, deoarece gândurile nu sunt observabile din exterior. În consecință, nu aveau de unde să știe ce te-a determinat să-ți faci un ceai.

Marele neurolog Jean-Martin Charcot a spus la un moment dat: „Teoria e bună, dar nu împiedică lucrurile să existe”[2]. Întrucât evenimentele mintale interne există în mod clar și influențează cauzal comportamentul, abordarea behavioristă a

fost eclipsată treptat în a doua jumătate a secolului XX de alta. Aceasta a fost numită psihologie „cognitivă”, termen ce include procesele psihice interioare, ca să spunem așa.

Stimulentul din spatele revoluției cognitive a fost apariția calculatoarelor. Behavioriștii considerau că funcționarea minții constituie o „cutie neagră” impenetrabilă, axându-se, în schimb, pe inputuri și outputuri. Dar calculatoarele nu sunt impenetrabile. Ne-ar fi fost imposibil să le inventăm fără să le înțelegem pe deplin funcționarea internă. De aceea, pentru că tratau mintea ca pe un calculator, psihologii s-au simțit încurajați să formuleze modele ale *prelucrării informațiilor* care se petrece în interiorul ei. Modelele lor au fost apoi testate cu ajutorul unor simulări artificiale ale proceselor psihice, în combinație cu experimente behavioriste.

Ce este prelucrarea informațiilor? Voi spune mai multe despre ea mai jos, dar cel mai interesant e că poate fi implementată cu tipuri foarte diferite de echipamente fizice. Astfel, reușim să vedem natura fizică a minții într-o lumină nouă. Ni se sugerează că mintea (concepută ca prelucrare a informațiilor) este o *funcție*, și nu o structură. Din această perspectivă, funcțiile de „software” ale minții sunt implementate de structurile „hardware” ale creierului, dar aceleași funcții pot fi implementate la fel de bine şi de alte substraturi, precum calculatoarele. Astfel, atât creierii, cât și calculatoarele îndeplinesc funcții de *memorie* (codează și stochează informații) și de *percepție* (clasifică tipare ale informațiilor de input prin compararea lor cu informațiile stocate), precum și funcții de *execuție* (execută decizii referitoare la ce este de făcut ca reacție la asemenea informații).

Aceasta este puterea a ceea ce a ajuns să se numească abordarea „funcționalistă”, dar este și punctul ei slab. Dacă aceleași funcții pot fi îndeplinite şi de calculatoare, despre care presupunem că nu sunt ființe conștiente, atunci se justifică să reducem mintea la simpla prelucrare a unor informații? Până și telefonul tău are funcții de memorie, de percepție și de execuție.

Al treilea răspuns științific important la metafizica minte/corp s-a dezvoltat în tandem cu psihologia cognitivă, dar, până la sfârşitul secolului trecut, a ajuns să o eclipseze. Mă refer la o abordare numită în general „neuroștiință cognitivă”. Aceasta se axează pe partea de hardware a minții și a apărut odată cu dezvoltarea a numeroase tehnici psihologice care ne permit să observăm și să măsurăm direct dinamica creierului viu.

Pe vremea behaviorismului, neurofiziologii erau limitați la o singură asemenea tehnică: puteau să înregistreze activitatea electrică a creierului de la suprafața exterioară a scalpului cu ajutorul unei electroencefalograme (EEG). Astăzi, avem mult mai multe instrumente la dispoziție, precum imagistica prin rezonanță magnetică funcțională (RMNf), pentru a măsura frecvența activității hemodinamice în diferite părți ale creierului în timp ce execută operaţii mintale specifice, și tomografia cu emisie de pozitroni (PET CT), cu care putem măsura activitatea metabolică diferențială pentru fiecare sistem de neurotransmițători. Astfel avem posibilitatea să identificăm cu precizie ce procese cerebrale ne generează diferitele stări psihice. De asemenea, avem posibilitatea să vizualizăm conectivitatea funcțional-anatomică detaliată între diferitele regiuni cerebrale folosind tractografia. Iar cu ajutorul

optogeneticii putem să vedem și să activăm circuitele de neuroni ce alcătuiesc urmele unor amintiri individuale, pe măsură ce luminează în timpul unor sarcini cognitive.

Aceste tehnici fac vizibilă clar funcționarea internă a organului minții – astfel concretizând cele mai hazardate visuri empiriste ale behavioriștilor, fără a limita sfera psihologiei la stimuli și răspunsuri.

Stadiul neuropsihologiei în anii 1980, când am intrat eu în acest domeniu, explică de ce behavioriștii au făcut o tranziție atât de ușoară de la teoria învățării la neuroștiința cognitivă. Neuropsihologia de la acea vreme putea fi numită la fel de bine neurobehaviorism. Cu cât ni se preda despre funcții precum memoria pe termen scurt, despre care se spunea că este un „amortizor" pentru reținerea amintirilor în conștiință, cu atât realizam că profesorii mei vorbeau despre altceva decât ceea ce urmăream eu să aflu. La cursuri ni se vorbea despre instrumentele funcționale folosite de minte, și nu despre minte în sine. Nu știam ce să mai înțeleg.

Neurologul Oliver Sacks, în cartea sa *Un picior de sprijin* (1984), a descris sugestiv situația în care mă aflam:

> Neuropsihologia, ca şi neurologia clasică, urmărește să fie cu totul obiectivă, iar marea sa forță, progresele sale tocmai de aici vin. Dar o ființă vie, mai ales o ființă umană, este în primul rând activă – subiect, nu obiect. Tocmai subiectul, „eul" viu, este exclus. Neuropsihologia e admirabilă, dar exclude psihicul – exclude „eul" viu, activ, ce are experiențe[3].

Afirmația „Neuropsihologia e admirabilă, dar exclude psihicul” mi-a surprins perfect dezamăgirea. După ce am citit-o, am inițiat o corespondență cu Oliver Sacks care a continuat până la moartea sa, în 2015. M-a atras la el faptul că lua atât de în serios relatările subiective ale pacienților lui. Acest lucru era deja evident în cartea sa din 1970, *Migraine*, și chiar mai mult în extraordinarul său volum *Revenirea la viaţă* (1973). Cea de-a doua carte consemna cu detalii de mare finețe parcursul clinic al unui grup de pacienți cronici cu achinezie, mutism și encefalită letargică. Această afecțiune era numită și „boala somnului”, deși pacienții nu dormeau propriu-zis, ci doar nu manifestau inițiativă sau pulsiune spontană. Sacks „i-a trezit” administrându-le levodopă, o substanță care crește cantitatea de dopamină. Totuși, după ce au revenit la o intenționalitate activă, au devenit rapid excesiv de energici, maniacali și, în cele din urmă, psihotici. La scurt timp după ce am citit *Un picior de sprijin*, care descrie experiența subiectivă a unei leziuni a sistemului nervos trăită chiar de Sacks, acesta a publicat *Omul care îşi confunda soţia cu o pălărie* (1985) – o serie de studii de caz ce furnizează introspecții revelatoare legate de diverse afecțiuni neuropsihologice din perspectiva unor pacienți cu antecedente neurologice. Astfel Sacks a dobândit o faimă de durată.

Aceste cărți nu prea semănau cu manualele mele de neuropsihologie, ce disecau funcțiile mintale în modul în care am proceda în cazul funcțiilor oricărui organ. De exemplu, am aflat că limbajul este produs în aria Broca din lobul frontal stâng, că înțelegerea vorbirii are loc în aria Wernicke, la câțiva centimetri mai în spate, în lobul temporal, și că abilitatea de a repeta ce ți se spune este mediată de fasciculul arcuat, un tract

fibros ce conectează aceste două arii. De asemenea, am aflat că amintirile sunt codificate de hipocamp, stocate în neocortex și accesate prin mecanisme frontale-limbice.

Chiar nu este creierul diferit de stomac și de plămâni? Aspectul evident care îl individualizează e faptul că există „ceva cum e" *să fii* un creier. Acest lucru nu se aplică și altor părți ale corpului. Senzațiile pe care le localizăm în alte organe nu sunt simțite de acestea, ci impulsurile nervoase care pornesc din ele sunt simțite numai când ajung la creier. Cu siguranță exista un motiv pentru care țesutul cerebral are această proprietate foarte specifică – capacitatea de a percepe, a simți și a gândi lucruri. Această proprietate pare că face ceva. Și, dacă face – dacă experiențele subiective au efecte cauzale asupra comportamentului, așa cum pare când te hotărăști spontan să-ți faci un ceai –, atunci am fi într-o mare eroare dacă am omite-o din descrierile științifice. Totuși, exact asta se întâmpla în anii 1980. Profesorii nu ne-au spus nimic niciodată despre cum e să înțelegi vorbirea sau să accesezi o amintire, ca să nu mai vorbim de motivul pentru care simțim ce simțim.

Cei care ţineau seama de perspectiva subiectivă nu erau luați în serios de specialiștii în neuroștiințe. Nu știu câți oameni cunosc faptul că lucrările lui Sacks erau în mare parte ridiculizate de colegii lui. Unul dintre comentatori chiar l-a numit „Omul care și-a confundat pacienții cu o carieră literară". Toate acestea l-au afectat în mare măsură. Cum să descrii viața interioară a ființelor umane fără să le spui poveștile? După cum se lamentase Freud cu un secol înainte în privința rapoartelor sale clinice:

Continuă să mi se pară ciudat faptul că uneori istoricul bolii

scris de mine poate fi citit ca o nuvelă și că aceasta este lipsită, ca să spunem așa, de pecetea serioasă a travaliului științific. Trebuie să mă consolez cu faptul că, evident, pentru acest rezultat poate fi făcută răspunzătoare mai degrabă natura obiectului decât preferința mea[*1] [4].

Sacks a fost încântat când i-am trimis acest citat[5]. În ce mă privește, când am citit aceste rânduri pentru prima dată, mi-am dat seama că nu sunt singurul care a intrat în domeniul neuropsihologiei cu speranța că îmi va permite să învăț cum generează creierul subiectivitatea. Ajungi repede să nu mai crezi acest lucru. Ești avertizat să nu mai urmărești asemenea probleme de nerezolvat – ele sunt „nocive pentru cariera ta”. Și, astfel, majoritatea celor care studiază neuroștiințele uită treptat de ce au intrat în acest domeniu și ajung să se identifice cu dogma cognitivismului, care abordează creierul ca și cum nu ar fi deloc diferit de un telefon mobil.

Singurul aspect al conștiinței ce era într-adevăr o temă științifică respectabilă în anii 1980 era mecanismul cerebral al stării de veghe versus starea de somn. Cu alte cuvinte, „nivelul” de conștiință era un subiect respectabil, dar nu și „conținutul” ei. Așadar, m-am hotărât să-mi axez activitatea de cercetare doctorală pe un aspect al somnului. În particular, am ales să studiez aspectul subiectiv al somnului, şi anume mecanismele cerebrale ale visului. Până la urmă, visele nu sunt decât o imixtiune paradoxală a conștiinței („starea de veghe”) în starea de somn. În mod uimitor, exista o lacună semnificativă în literatura de specialitate pe acest subiect: nimeni nu descrisese sistematic cum anume degradarea unor părți ale creierului afectează visul. Așa că mi-am propus să o fac eu.

Visul este dificil de studiat tocmai din cauza naturii sale subiective. Fenomenele mintale în general nu pot fi observate decât introspectiv, de un singur observator, apoi raportate altora indirect, prin cuvinte. Dar visele sunt și mai problematice: nu pot fi raportate decât retrospectiv, după ce visul se încheie și cel care visează se trezește. Toată lumea știe că nu ne putem baza pe amintirile noastre despre vise. Ce fel de „date” sunt acestea[6]? Acesta este și motivul pentru care, începând de la mijlocul secolului XX, visele au constituit un front semnificativ în tranziția de la behaviorism la ceea ce avea apoi să devină neuroștiința cognitivă.

Electroencefalograma a fost aplicată pentru prima dată în studiul somnului la începutul anilor 1950 de doi neurofiziologi, Eugene Aserinsky și Nathaniel Kleitman. Ei au plecat de la ipoteza că nivelul activității cerebrale scade atunci când adormim și crește când ne trezim, prezicând de aici că amplitudinea undelor noastre cerebrale (unul dintre lucrurile măsurate prin electroencefalografie) crește, iar frecvența lor (celălalt lucru pe care îl măsoară) scade pe măsură ce adormim, precum și că se întâmplă invers când ne trezim (vezi figura 10, p. 136).

Când creierul intră în ceea ce numim acum somn „cu unde lente”, înțelegem exact ceea ce au prevăzut Aserinsky și Kleitman. Ipoteza lor s-a confirmat. Surpriza constă în ce se întâmplă în continuare: în decurs de aproximativ 90 de minute după ce adormim (apoi la fiecare circa 90 de minute, în cicluri regulate), undele cerebrale accelerează din nou, aproape atingând nivelul din starea de veghe, chiar dacă persoana de la care se obține înregistrarea rămâne în starea de somn[7]. Aserinsky și Kleitman au numit aceste etape ciudate ale

activării creierului „somn paradoxal” – paradoxul fiind că creierul este stimulat din punct de vedere fiziologic, deși doarme profund.

Se mai întâmplă tot felul de alte lucruri în această stare aparte. Ochii se mișcă rapid (motivul pentru care somnul paradoxal a fost ulterior denumit somn „cu mișcare rapidă a ochilor” sau somn REM), dar corpul de la gât în jos este paralizat temporar. Au loc și modificări involuntare spectaculoase, precum controlul redus al temperaturii bazale și mărirea organelor genitale, ducând la erecții vizibile la bărbați. Nu reușesc să înțeleg cum de nu a observat știința toate acestea până în 1953.

Pe baza acestor observații, Aserinsky și Kleitman au formulat o altă ipoteză, despre care nu se poate spune că nu era fondată: somnul REM este baza fiziologică a stării psihologice numite vis. În consecință, au prezis că, dacă suntem treziți din somnul REM, vom putea povesti vise, dar nu și dacă suntem treziți din somnul cu unde lente (non-REM). Împreună cu cercetătorul care a avut ghinionul să se numească William Dement, au testat această predicție și au confirmat-o: aproximativ 80% dintre trezirile din somnul REM și mai puțin de 10% din trezirile din somnul non-REM au produs evocarea unor vise. Din acel moment, somnul REM a fost considerat sinonim cu visul[8]. Ce veste excelentă! Domeniul nu mai trebuia să se preocupe de vis, deoarece acum avea un indicator obiectiv al acestuia, care le permitea specialiștilor în neuroștiințe să-și desfășoare activitatea ca la carte, fără a mai fi nevoiți să se confrunte cu complicațiile metodologice introduse de evocările verbale retrospective, cu un singur martor ale unor experiențe subiective efemere.

Mai exista un motiv de bucurie că scăpaserăm de vise. Acesta consta în rolul stânjenitor pe care îl jucaseră în consacrarea psihanalizei. Spre deosebire de răspunsurile clasice la metafizica minte/corp ce caracterizau știința minții în a doua jumătate a secolului XX, psihanaliștii nu aveau rețineri în a trata relatările introspective drept date. De fapt, relatările obținute prin „asociere liberă” (testarea nestructurată a fluxului conștiinței) constituiau principalele date ale studiilor psihanalitice. Folosind această metodă, Sigmund Freud a ajuns la concluzia că, în pofida aparenței absurde a unor experiențe onirice „manifeste”, conținutul lor „latent” (povestea de bază, dedusă de Freud din asocierile libere ale persoanei care visa) dezvăluia o funcție psihologică coerentă. Această funcție era *împlinirea dorințelor*.

Conform lui Freud, visul este ceea ce se întâmplă atunci când nevoile biologice care generează comportamentul din starea de veghe sunt dezinhibate în timpul somnului. Visele sunt încercări de satisfacere a acestor nevoi, care continuă să ne solicite chiar și în timpul somnului. Totuși, visele fac acest lucru într-un mod halucinatoriu, permițându-ne astfel să rămânem adormiți (în loc să ne trezim ca să ne satisfacem cu adevărat pulsiunile). Întrucât halucinațiile sunt o caracteristică esențială a tulburărilor mintale, Freud a folosit această teorie în lucrarea sa de referință *Interpretarea viselor* (1900) ca să descrie un model general al modului de funcționare a minții în ansamblul ei, în condiții de sănătate și de boală.

După cum a spus Freud, „psihanaliza se bazează pe analiza visului”[9]. Dar visele, după cum am văzut, sunt incredibil de dificil de studiat în mod empiric, așa că behavioriștii le-au eliminat din știință. În plus, edificiul teoretic construit de Freud

pe vise nu era mai bun decât fundamentele sale. Marele filosof al științei Karl Popper a declarat teoria psihanalitică drept „pseudoștiințifică”, din cauză că nu genera predicții falsificabile din punct de vedere experimental[10]. Cum falsifici ipoteza dedusă de Freud, conform căreia visele exprimă dorințe latente? Dacă nu e neapărat ca dorințele să apară în visul manifest (relatat), atunci orice vis poate fi „interpretat” pentru a se potrivi cerințelor teoriei. Așadar, deloc surprinzător, când descoperirea somnului REM le-a permis specialiștilor în neuroștiințe să treacă de la materia efemeră a evocărilor viselor la corelativele lor fiziologice concrete, visele le-au scăpat printre degete.

Descoperirea somnului REM în anii 1950 a declanșat o cursă pentru identificarea bazei sale neurologice, întrucât funcționarea somnului REM putea dezvălui mecanismul *obiectiv* al viselor, a cărui elucidare putea să plaseze psihiatria vremii pe o poziție științifică mai respectabilă. (Această cercetare a fost simplificată datorită faptului că somnul REM apare la toate mamiferele). Cursa a fost câștigată de Michel Jouvet în 1965. Într-o serie de experimente chirurgicale pe pisici, Jouvet a demonstrat că somnul REM era generat nu de creierul anterior (ce include cortexul, partea exterioară a creierului care este atât de impresionant de mare la oameni și, parțial din acest motiv, este considerată organul minții), ci de trunchiul cerebral, o structură presupusă a fi mult mai modestă, de o origine evolutivă extrem de veche[11]. Jouvet a ajuns la această concluzie observând că secțiunile progresive de la nivelul creierului, începând din partea superioară și continuând în jos, produceau pierderea somnului REM numai atunci când secționarea ajungea la nivelul unei structuri

„inferioare” a trunchiului cerebral numită punte (vezi figura 1)[12].

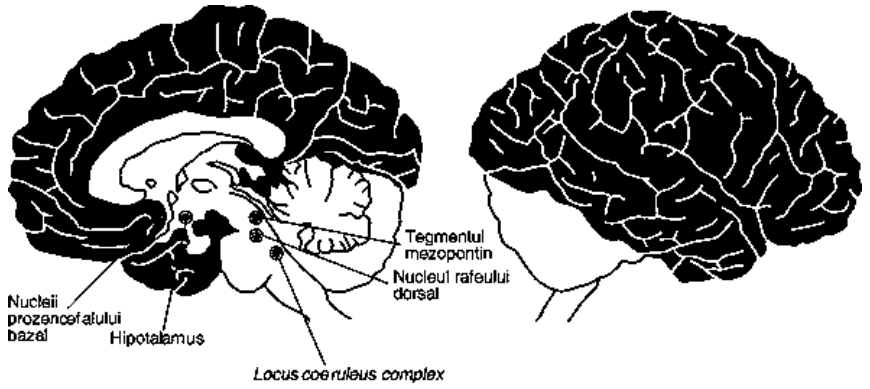


Figura 1. Imaginea din stânga este o reprezentare mediană a creierului (secționat prin mijloc), iar cea din dreapta este o reprezentare laterală (văzut dintr-o parte). Figura arată cortexul (negru) și trunchiul cerebral (alb). Sunt indicați numai nucleii trunchiului cerebral considerați importanți pentru controlul somnului REM, şi anume tegmentul mezopontin, nucleul rafeului dorsal și *locus coeruleus complex*. Este reprezentată și localizarea nucleilor prozencefalului bazal (sub cortex) și a hipotalamusului, a cărei relevanță se va vedea mai jos.

Sarcina de a se ocupa de detalii i-a revenit lui Allan Hobson, un student al lui Jouvet. Hobson a identificat exact grupările de neuroni pontini care generau somnul REM și, prin urmare, visele. Până pe la mijlocul anilor 1970 a devenit evident că întregul ciclu somn/stare de veghe – inclusiv toate fenomenele somnului REM enumerate mai sus, precum și cele din diferitele etape ale somnului non-REM – era orchestrat de un număr mic

de nuclei din trunchiul cerebral care interacționau între ei[13]. Cei care controlează somnul REM semănau cu un întrerupător simplu de tip pornit/oprit. Neuronii care *declanșează* somnul REM se află în tegmentul mezopontin (vezi figura 1). Aceștia secretă în tot creierul anterior o substanţă neurochimică numită acetilcolină. Acetilcolina provoacă activarea: crește „nivelul” conștiinței (de exemplu, este stimulată de nicotină, ceea ce te ajută să te concentrezi). Neuronii din trunchiul cerebral care *opresc* somnul REM se află mai în profunzimea punții, în rafeul dorsal și în *locus coeruleus complex* (din nou, vezi figura 1). Ei eliberează serotonină și, respectiv, noradrenalină. La fel ca acetilcolina, acești compuși neurochimici modulează diferite aspecte ale nivelului de conștiință.

Combinând aceste concluzii cu faptul că somnul REM se declanșează și se oprește automat, cam la fiecare 90 de minute, ca un ceas, Hobson a conchis inevitabil: „Principala forță ce declanșează visele nu este de natură psihologică, ci fiziologică, deoarece momentul apariției și durata somnului cu vise sunt destul de constante, sugerând o geneză preprogramată, determinată neuronal”[14].

Deoarece somnul REM apare din trunchiul cerebral colinergic, o parte străveche și inferioară a creierului departe de maiestuosul cortex unde se presupune că are loc toată acțiunea psihologiei umane, Hobson a adăugat că visul nu poate fi stimulat de dorințe; e „neutru din punctul de vedere al stimulării”[15]. Prin urmare, conform lui Hobson, opinia lui Freud că visele vin din dorințe latente este probabil complet greșită. Semnificația pe care Freud a văzut-o în vise nu le era mai intrinsecă decât le este petelor de cerneală. Aceasta era

proiectată asupra lor, neexistând în visul în sine. Din punct de vedere științific, interpretarea viselor nu era cu nimic mai bună decât ghicitul în frunze de ceai.

Deoarece întreaga psihanaliză era fundamentată pe metoda folosită de Freud pentru a studia visele, toate teoriile derivate astfel de acesta puteau fi respinse. Ca urmare a demontării de către Hobson a ideii că visele ar avea vreo valoare, psihiatria putea, în sfârșit, să se îndepărteze de fundamentarea ei istorică pe relatările introspective și să se bazeze de-acum pe metodele obiective ale neuroștiințelor (mai ales neurochimia) pentru cercetare și tratament. În consecință, dacă în anii 1950 era aproape imposibil să devii profesor titular de psihiatrie la o universitate americană importantă fără să fii şi psihanalist, astăzi e invers: este aproape imposibil să devii profesor de psihiatrie dacă *ești* psihanalist.

Nimic din toate acestea nu m-a surprins în mod deosebit la vremea respectivă. Problema din centrul cercetărilor mele doctorale părea destul de clară și deloc implicată în luptele pentru moștenirea freudismului și a behaviorismului. Eu nu voiam să știu decât cum este afectată experiența reală a visului de leziunile la nivelul diferitelor părți ale creierului anterior și ale cortexului său. Până la urmă, dacă acțiunea avea loc în prozencefal, din punct de vedere psihologic, cu siguranță că trebuia să facă *ceva* în vise.

Departamentul de neurochirurgie din cadrul University of Witwatersrand avea secții în două spitale universitare – Baragwanath Hospital și Johannesburg General Hospital. Baragwanath era un fost spital militar mare, în complexul urban „noneuropean” Soweto. Având în vedere că toate acestea

se întâmplau la apogeul apartheidului, era un ocean de suferință umană. În schimb, Johannesburg General Hospital, rezervat „europenilor”, era un spital universitar modern, un monument închinat inegalității rasiale. De asemenea, departamentul de neurochirurgie avea paturi și în secția de recuperare după leziuni cerebrale și spinale din cadrul Edenvale General Hospital, care funcționa într-o veche clădire colonială dintr-o suburbie a Johannesburgului. Începând din 1985, am lucrat în toate aceste trei locuri, examinând sute de pacienți anual. I-am inclus pe 361 dintre ei în studiile mele doctorale, care s-au desfășurat în următorii cinci ani.

După ce am învățat să folosesc tehnologia electroencefalografiei și altele conexe și să recunosc undele cerebrale caracteristice asociate cu diferitele etape ale somnului, am reușit să-i trezesc pe subiecți în timpul somnului REM, când era cea mai mare probabilitate să viseze. De asemenea, am chestionat pacienți cu antecedente neurologice aflați la pat despre modificări ale viselor lor, apoi i-am monitorizat zile, săptămâni și luni la rând. Astfel am investigat dacă conținutul viselor era afectat sistematic de leziunile localizate în diferite arii cerebrale. În pofida reputației îndoielnice a relatărilor viselor, am presupus că, dacă pacienții cu leziuni la nivelul aceleiași arii cerebrale reclamau aceeași modificare a conținutului viselor, aveam toate motivele să-i cred. Această metodă se numește „corelare clinico-anatomică”: prin examinarea clinică a abilităților psihologice ale pacienților, observi cum se modifică o funcție cerebrală ca urmare a unei leziuni la nivelul unei anumite părți a creierului; apoi corelezi această modificare cu locul leziunii, astfel descoperind indicii cu privire la funcția structurii cerebrale

lezate, ceea ce duce la ipoteze testabile. Această metodă se aplica sistematic de multe decenii în cazul tuturor funcțiilor cognitive principale, precum percepția, memoria și limbajul, dar încă nu fusese aplicată în cazul visului.

La început, am fost puțin stânjenit să vorbesc cu oameni atât de grav bolnavi despre visele lor. Pe mulți dintre ei îi aștepta sau tocmai trecuseră printr-o intervenție chirurgicală la nivelul creierului ce le putea pune viața în pericol și, în asemenea condiții, mă temeam să nu îmi considere întrebările frivole. Dar pacienții mei erau surprinzător de dispuși să descrie modificările din viața lor psihică generate de afecțiunile neurologice.

Până să-mi încep cercetările, fuseseră publicate mai multe studii de caz în care efectul observat în experimentele pe animale apărea şi la oameni, şi anume că somnul REM nu mai survenea dacă existau leziuni la nivelul tegmentului mezopontin (vezi figura 1). M-a surprins însă că nu se gândise nimeni să pună întrebări despre modificările survenite în *visele* pacienților. Este un exemplu cât se poate de clar al ideilor preconcepute din neuroștiințe față de datele subiective[16].

În cercetarea mea, m-am așteptat să descopăr ceea ce era evident: că pacienții cu leziuni la nivelul cortexului vizual au vise nonvizuale, că pacienții cu leziuni la nivelul cortexului responsabil de limbaj au vise nonverbale, că pacienții cu leziuni la nivelul cortexului somatosenzorial și motor au vise hemiplegice și așa mai departe. Acestea sunt ideile de bază în ce privește corelația creier-comportament. Acesta era golul pe care voiam să-l umplu și am făcut-o bucuros[17].

Totuși, spre uimirea mea, pe lângă toate chestiunile evidente pe care le-am observat, am descoperit și că pacienții cu leziuni

la nivelul ariei cerebrale care generează somnul REM *încă aveau vise*. Mai mult, pacienții care *nu* mai visau suferiseră leziuni în arii cerebrale complet diferite. Prin urmare, visul și somnul REM erau ceea ce numim fenomene „dublu disociabile”[18]. Se corelau reciproc (adică de obicei surveneau în același timp), dar nu erau același lucru[19].

Timp de aproape 50 de ani, în domeniul științei somnului, cei care studiau creierul au confundat corelația cu identitatea. De îndată ce au stabilit că visul însoțește somnul REM, s-au grăbit să tragă concluzia că sunt unul și același lucru – apoi s-au debarasat de partea subiectivă problematică a corelației. Din acel moment, cu foarte puține excepții, au studiat exclusiv somnul REM, în principal în cadrul unor experimente pe animale, care nu pot furniza relatări introspective. Eroarea a ieșit la suprafață abia atunci când am început să studiez din perspectiva neuroștiințelor *experiența* viselor la pacienții cu antecendente neurologice.

La începutul anilor 1990, când am semnalat pentru prima dată că visele dispar odată cu leziunile la nivelul altei arii cerebrale decât cea care generează somnul REM, am insistat să subliniez faptul că aria critică nu se află în trunchiul cerebral[20]. Am procedat astfel deoarece voiam să evidențiez natura mintală a viselor, și știam toți că sediul funcțiilor mintale este cortexul.

De fapt, am identificat două regiuni în care leziunile duc la pierderea viselor cu păstrarea somnului REM. Prima se află în cortex, în lobulul parietal inferior (vezi figura 2). Această descoperire nu a fost surprinzătoare, întrucât lobul parietal este important pentru memoria pe termen scurt. Dacă un pacient nu își poate menține conținutul memoriei în

amortizorul conștiinței, cum să viseze? Mult mai interesantă a fost a doua regiune cerebrală, şi anume substanța albă din cadranul ventromedial al lobilor frontali, care conectează cortexul frontal cu diferite structuri subcorticale. Această descoperire a fost cu totul neașteptată. Nimic din funcțiile acestei regiuni cerebrale nu are o conexiune evidentă cu experiența manifestă a visului, dar cu siguranță contribuie cu ceva esențial la acest proces, întrucât leziunile la nivelul acestei regiuni au cauzat în mod sigur dispariția totală a viselor.

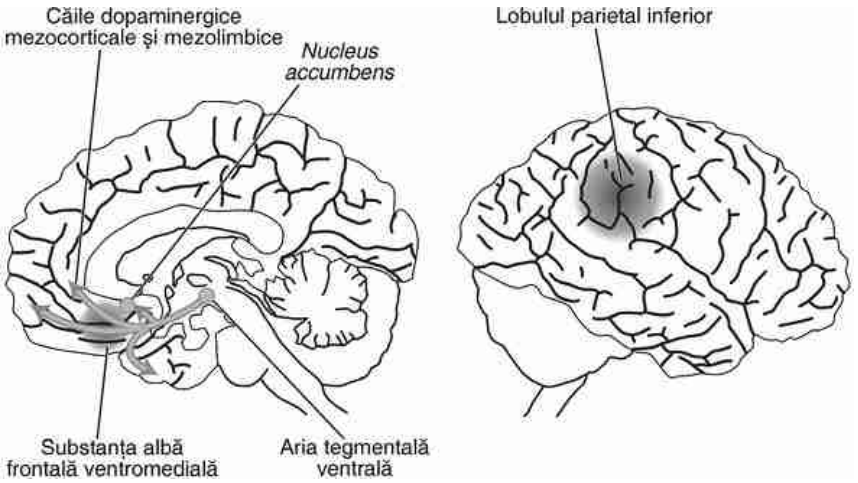


Figura 2. Cele două regiuni cu leziuni care au dus la oprirea viselor sunt reprezentate cu o umbră în această figură: substanța albă ventromedială din lobul frontal (stânga) și lobulul parietal inferior al cortexului (dreapta). Sunt reprezentate și aria tegmentală ventrală a trunchiului cerebral și principalele căi fibroase ce pornesc de acolo, şi anume căile dopaminergice mezocorticale şi mezolimbice. Notă: locul leziunii la nivelul lobului frontal ventromedial implică aceste căi subcorticale, care trec pe sub cortex, și nu prin el. O destinație importantă a acestor căi este *nucleus accumbens*, de asemenea reprezentat.

Spun „în mod sigur”, deși am raportat doar nouă cazuri de pierdere a viselor în rândul pacienților mei cu leziuni ale lobului frontal (și 44 de cazuri pentru leziunile parietale). Asemenea leziuni sunt extrem de rare în practica clinică obișnuită. Totuși, corelația este sigură. În prima jumătate a secolului XX, substanța albă frontală ventromedială a fost vizată chirurgical în mii de cazuri printr-o tehnică numită leucotomie prefrontală modificată. Psihiatrii acelor zile pline de entuziasm au descoperit că unele afecțiuni mintale grave puteau fi ameliorate prin distrugerea chirurgicală completă a lobilor prefrontali (cunoscută sub denumirea tehnică de lobotomie frontală), dar au observat și că această procedură radicală are numeroase „efecte secundare”, în termenii lor eufemistici. Prin urmare, au redus amploarea leziunii, încercând să identifice care este cea mai mică parte a lobilor frontali ce poate fi deconectată de restul creierului pentru a se obține totuși rezultatele dorite. Procedura modificată de Walter Freeman și James Watts a fost soluţia. Aceasta presupunea inserarea unui bisturiu mic cu vibrații prin orbitele ochilor, care secționa substanța albă din cadranul ventromedial al lobilor frontali (leucotomie prefrontală), exact în locul leziunilor celor nouă pacienți ai mei.

Așadar, m-am întors la vechea literatură despre psihochirurgie ca să văd dacă îmi confirmă ce am observat la cazurile mele[21]. Aveam motive să sper că doctorii care au examinat pacienții cu leucotomie clasică i-au întrebat despre vise după intervenții – până la urmă, psihiatrii încă luau visele în serios pe atunci. S-a dovedit că am avut dreptate. Aceștia au

descoperit că leucotomia prefrontală avea trei efecte psihologice principale. În primul rând, reducea simptomele psihotice pozitive (halucinațiile și delirurile). În al doilea rând, reducea stimularea. În al treilea rând, cauza pierderea viselor. De fapt, unul dintre primii cercetători în domeniul psihochirurgiei a sugerat chiar că menținerea capacității de a visa după intervenție era un semn de prognostic nefavorabil[22].

Această ultimă idee m-a ajutat să intuiesc care dintre numeroasele circuite neuronale din cadranul ventromedial al lobilor frontali era cel mai probabil să fie responsabil pentru pierderea viselor. De asemenea, mi-a furnizat un prim indiciu în ce privește motivul pentru care trebuie să căutăm vinovatul în această regiune cerebrală la care nu ne-am aștepta. Ce sunt visele dacă nu halucinații și deliruri? Iată de ce ar fi un prognostic nefavorabil dacă acestea ar persista după leucotomie.

Tratamentul neurochirurgical al halucinațiilor și delirurilor nu a fost abandonat din rațiuni etice, ci a căzut în dizgrație când a început să se vadă că se pot obține rezultate terapeutice echivalente cu o rată de morbiditate și mortalitate mai mică folosind medicamente devenite disponibile în masă prima dată în anii 1950, și anume „tranchilizantele majore”. Aceste medicamente și „antipsihoticele” moderne blochează dopamina neurochimică la capetele unui circuit cerebral numit sistemul dopaminergic mezocortical şi mezolimbic (vezi figura 2). Întrucât acest circuit este întrerupt prin leucotomia prefrontală, cum s-a întâmplat la cei nouă pacienți ai mei cu leziuni survenite natural, am presupus că acesta ar putea fi sistemul ce generează visele.

Alte experimente mi-au confirmat ipoteza. Se stabilise deja

că stimularea farmacologică a acestui circuit creștea frecvența, durata și intensitatea viselor, fără efecte proporționale asupra somnului REM[23]. Era vorba de levodopă, medicamentul folosit de Oliver Sacks pentru a-și „trezi” pacienții postencefalie. Neurologii care folosesc stimulenți ai dopaminei în tratamentul bolii Parkinson știu de multă vreme că trebuie să fie atenți să nu își împingă pacienții spre psihoză, cum a făcut Sacks, iar apariția unor vise neobișnuit de intense este adesea primul semn al acestui efect secundar[24]. Observațiile ulterioare cruciale au fost că neuronii ce alcătuiesc acest circuit (ai cărui corpi celulari se găsesc în aria tegmentală ventrală) se descarcă la o rată maximă în timpul somnului cu vise[25] și, în același timp, transportă dopamină în cantități maxime către țintele lor din *nucleus accumbens* (vezi figura 2)[26]. Așadar, astăzi se consideră în general că visele pot să apară independent de somnul REM și că circuitul dopaminergic mezocortical-mezolimbic este într-adevăr principalul stimulent al viselor[27].

Leziunile la nivelul căilor colinergice din cadranul ventromedial al lobilor frontali (care apar din nucleii prozencefalului bazal, vezi figura 1) produc efectul opus faţă de ceea ce se întâmplă când există leziuni la nivelul căilor dopaminergice, şi anume *mai multe* vise, nu mai puține. Hobson afirmase că acetilcolina era generatorul viselor neutru din punctul de vedere al stimulării, dar același lucru se întâmplă şi când blochezi acetilcolina pe cale farmacologică, ca atunci când apar leziuni la nivelul căilor. Acum se știe despre medicamentele anticolinergice – blocanții acetilcolinei – că provoacă vise *în exces*[28]. Cu alte cuvinte, blocarea sistemului neuronal despre care Hobson spunea că este responsabil pentru apariția viselor are efectul opus celui prognozat de

teoria lui.

În scurt timp, a fost clar că neuroștiințele îi datorau scuze lui Freud. Dacă există o parte a creierului care poate fi considerată responsabilă cu „dorințele”, aceasta este circuitul dopaminergic mezocortical-mezolimbic. Este orice numai stimulent neutru nu. Edmund Rolls (și mulți alții) numește acest circuit sistemul de „recompensare” al creierului[29]. Kent Berridge îl numește sistemul „dorințelor”. Jaak Panksepp îl numește sistemul CĂUTĂRII – și atrage atenția asupra rolului său pentru funcția de *căutare a hranei*[30]. Acesta este circuitul cerebral responsabil cu „comportamentele de explorare și de căutare cele mai energice pe care le poate manifesta un animal”[31]. Este și circuitul care stimulează visele[32].

Hobson nu a fost amuzat. M-a invitat să-mi prezint concluziile în fața grupului său de cercetare din cadrul departamentului de neurofiziologie de la Harvard. Inițial, le-a acceptat și a publicat o recenzie favorabilă a cărții pe care am scris-o pe acest subiect în 1997, notând că rezultatele mele clinico-anatomice au fost confirmate până la ultimul detaliu de studiile de neuroimagistică ale lui Allen Braun (vezi figura 3, p. 47)[33]. Apoi și-a dat seama că aceste evoluții ar putea reabilita o viziune în mare parte freudiană asupra viselor, punct în care mi-a scris că e dispus să-mi sprijine public concluziile cu condiția să nu pretind că îi dau dreptate lui Freud. Cam asta se poate spune despre presupusa obiectivitate a neuropsihologiei.

Totuși, descoperirile mele mai aveau un aspect foarte surprinzător. Când am dat peste el pentru prima dată, nu am acordat mare atenție faptului că neuronii care activează acest circuit sunt localizați în *trunchiul cerebral* (la fel ca şi cei din circuitele care generează somnul REM). După cum am spus, am

vrut să evidențiez natura mintală a viselor. A fost nevoie ca Allen Braun, specialistul în neuroimagistică pe care tocmai l-am menționat, să-mi atragă atenția politicos asupra omisiunii mele. În contextul dezacordului științific dintre mine și Hobson în privința circuitelor cerebrale care activează procesul visului (dopaminergice sau colinergice), Braun a scris:

> Ciudat este că, după ce argumentează că structurile prozencefalului joacă un rol esențial în sistemul visului, Solms ajunge să sugereze că circuitele dopaminergice aferente acestor regiuni generează visele – *prin aceasta plasând declanșatorul viselor înapoi în trunchiul cerebral*[34].

Braun a concluzionat: „Mi se pare că acești domni încep să ajungă la un consens”[35]. În anii 1990, credeam, la fel ca restul neuropsihologilor, că toată activitatea psihologică avea loc în cortex, așa că m-am concentrat asupra faptului că tracturile substanței albe care mă interesau erau în lobii frontali, unde se localizau leziunile în cele nouă cazuri ale mele. Dar toți nucleii trunchiului cerebral au axoni lungi orientați în sus, spre prozencefal (vezi figura 2). Corpii celulari ai acestor neuroni sunt localizaţi în trunchiul cerebral, deși fibrele care pornesc din ei (axonii) se termină în cortex. Acest lucru susține principala funcție de *activare* a acestor nuclei ai trunchiului cerebral, care împreună se numesc sistemul reticulat activator. Tocmai aceste căi activatoare erau lezate în cazul celor nouă pacienți ai mei și al sutelor de pacienți supuși leucotomiei documentați anterior care nu mai visau.

Începând din 1998, impulsionat şi de comentariile lui Braun despre implicațiile descoperirii mele, mi-am îndreptat atenția spre celelalte sisteme de activare din trunchiul cerebral.

Studiile cele mai interesante în acest domeniu le desfășura Jaak Panksepp, a cărui lucrare enciclopedică *Affective Neuroscience* (1998) prezintă minuțios tot felul de dovezi în sprijinul opiniei sale că aceste sisteme presupuse a fi independente de conștiință, responsabile pentru reglarea doar a „nivelului” de conștiință, generează un „conținut” propriu.

Acest lucru s-a dovedit a fi foarte semnificativ.

*1 Sigmund Freud, *Studii despre isterie*, traducere din limba germană de Maria și Ion Nastasia, cuvânt înainte de Vasile Dem. Zamfirescu. București: Trei, 2014, p. 242 (n.tr.).

2

Înainte de Freud și după

În 1987 am mai luat o decizie care m-a pus în dezacord cu restul colegilor mei. M-am hotărât să mă specializez ca psihanalist[1]. Concluziile ce se conturau din studiile mele asupra viselor mă convinseseră că relatările subiective joacă un rol vital în neuropsihologie și că opoziția față de Freud dusese domeniul pe mai multe căi greșite. Dar nu rezultatele cercetărilor mele au fost factorul decisiv.

Ce m-a făcut să mă hotărăsc a fost un seminar la care am participat la University of the Witwatersrand, pe la mijlocul anilor 1980, coordonat de un profesor de literatură comparată pe nume Jean-Pierre de la Porte. Seminarul se referea la *Interpretarea viselor*, care îmi stârnise curiozitatea în contextul cercetării mele pentru doctorat. Ca toată lumea în acea perioadă, eram sceptic cu privire la Freud. Învățasem din studenție că psihanaliza este o „pseudoștiință”. Nimeni din domeniul științelor exacte nu îl mai lua în serios pe Freud, acesta fiind probabil motivul pentru care seminarul se desfășura în cadrul unui departament de științe umaniste. Eu am participat pentru că Freud fusese dispus să vorbească despre *conținutul* viselor, tema mea de cercetare.

De la Porte a explicat că nu putem înțelege concluziile teoretice la care a ajuns Freud fără să-i analizăm mai întâi un manuscris anterior, scris în 1895, dar publicat abia în anii 1950, după moartea sa. Manuscrisul se intitulează „Proiect de psihologie”[2] şi în el Freud a încercat să-și prezinte primele idei despre minte din perspectiva neuroștiințelor.

Astfel, călca pe urmele marelui său profesor, fiziologul Ernst von Brücke, membru fondator al Societății de Fizică din Berlin. Misiunea acestei societăți a fost formulată de Emil du Bois-Reymond în 1842 după cum urmează:

> Eu și Brücke am jurat solemn să traducem în fapt acest adevăr: „În organism nu acționează nicio altă forță în afară de cele fizice și chimice obișnuite. În cazurile ce nu pot fi explicate actualmente prin aceste forțe, trebuie fie să identificăm calea sau forma specifică a acțiunii lor prin metoda fizico-matematică, fie să presupunem că există forțe noi, egale ca importanță cu forțele chimico-fizice inerente materiei, reductibile la forța de atracție și de respingere”[3].

Iubitul lor profesor, Johannes Müller, se întrebase cum și de ce viața organică diferă de materia anorganică. A concluzionat că „organismele vii sunt fundamental diferite de entitățile nonvii deoarece conțin un oarecare element nonfizic sau sunt guvernate de alte principii decât lucrurile neînsuflețite”[4]. Pe scurt, conform lui Müller, organismele vii au o „energie vitală” sau „o forță vitală” pe care legile fiziologice nu o pot explica. El considera că ființele vii nu pot fi reduse la mecanismele lor fiziologice componente deoarece sunt unități indivizibile cu *obiective* și *scopuri*, pe care Müller le atribuia faptului că ele au

un suflet. Dat fiind că termenul german *Seele* poate fi tradus fie prin „suflet”, fie prin „minte”[5], dezacordul dintre Müller și studenții lui seamănă izbitor de mult cu dezbaterea aprinsă din zilele noastre dintre filosofi ca Thomas Nagel și Daniel Dennett cu privire la întrebarea dacă *conștiința* poate fi redusă la niște legi fizice (Nagel pretinde că nu, Dennett pretinde că da).

Când am asistat la seminarul lui de la Porte, surprizător pentru mine a fost să aflu că Freud – deschizătorul de drumuri în ce privește investigarea subiectivității umane – se aliniase nu cu vitalismul lui Müller, ci cu fizicalismul lui Brücke. După cum a scris în primele rânduri ale „Proiectului” său din 1895: „[Este] intenția de a oferi o psihologie științifică, adică de a prezenta procesele psihice ca pe niște stări determinate cantitativ ale unor particule materiale detectabile”[*1] [6].

Nu realizasem că Freud gândea în termenii neuroștiințelor. Acum aflam că abandonase cu reticență metodele neurologice de cercetare când a înțeles, cândva între 1895 și 1900, că metodele disponibile pe atunci nu erau capabile să expună fundamentele fiziologice ale minții.

Totuși, schimbarea de atitudine a lui Freud a avut multe părți bune. L-a forțat să analizeze mai atent fenomenele psihologice ca atare și să elucideze mecanismele funcționale aflate la baza lor. Astfel a apărut modul psihologic de cercetare pe care l-a numit „psihanaliză”. Premisa fundamentală a acestuia era că fenomenele subiective manifeste (numite astăzi „explicite” sau „declarative”) au cauze latente (numite astăzi „implicite” sau „nondeclarative”). Adică Freud susținea că fluxul haotic al gândurilor noastre conștiente poate fi explicat numai dacă presupunem că există conexiuni implicite între ele de care nu suntem conștienți. De aici au apărut noțiunea de

funcții mintale latente și celebra conjectură a lui Freud despre intenționalitatea „inconștientă”.

Întrucât la sfârșitul secolului al XIX-lea nu existau metode de investigare a fiziologiei evenimentelor mintale inconștiente, mecanismele lor puteau fi deduse doar prin observație clinică. Ceea ce a aflat Freud astfel a generat a doua sa afirmație fundamentală. A observat că pacienții adoptau o atitudine nici pe departe indiferentă față de intențiile lor inconștiente deduse – mai degrabă, ei păreau că *nu vor* să le conștientizeze, și nu că ar fi incapabili. Freud a folosit diverse denumiri pentru această tendință – „rezistență”, „cenzură”, „apărare” și „reprimare” – și a observat că previne detresa emoțională. Acest lucru, la rândul său, a relevat rolul central jucat de *sentimente* în viața psihică, modul în care susțin ele tot felul de biasuri cognitive. Aceste concluzii (evidente astăzi) i-au arătat lui Freud că unele dintre principalele forțe ce stimulează viața psihică sunt în întregime subiective, dar și inconștiente. Investigarea lor sistematică l-a condus la a treia sa afirmație fundamentală. Freud a concluzionat că în spatele sentimentelor stau în primul rând nevoile corpului, că, și la oameni, și la animale, viața psihică este *impulsionată* de imperativele biologice ale supraviețuirii și reproducerii. Aceste imperative, considera Freud, constituie legătura dintre mintea care simte și corpul fizic.

Freud a abordat într-un mod remarcabil de subtil relația minte-corp. A realizat că fenomenele psihologice pe care le studia nu erau direct reductibile la cele fiziologice. Încă din 1891 afirma că nu e posibil să atribui simptome psihologice unor procese neurofiziologice înainte de a reduce fenomenele psihologice și fiziologice relevante (ambele părți ale ecuației) la *funcțiile* lor de bază. După cum am observat mai sus cu privire

la prelucrarea informațiilor, funcțiile pot fi executate la diverse niveluri[7]. Freud considera că psihologia și fiziologia pot fi reconciliate numai pe terenul comun al funcției. Obiectivul său era să explice fenomenele psihologice prin intermediul unor legi funcționale „metapsihologice” (termenul înseamnă „dincolo de psihologie”)[8]. Încercarea de a sări peste acest nivel funcțional al analizei, de a trece direct de la psihologie la fiziologie, se numește astăzi eroarea localizaționistă[9].

În mod clar, pentru Freud, dacă nu și pentru adepții lui, psihanaliza trebuia să fie o etapă intermediară. Deși, de la bun început, el a căutat să identifice legile care ne guvernează viața interioară bogată în experiențe subiective, viața psihică a rămas pentru el o problemă *biologică*[10]. După cum a scris în 1914, „toate iterimatele noastre psihologice trebuie aşezate odată pe solul purtătorului organic”[*2] [11]. Prin urmare, a anticipat cu entuziasm ziua în care psihanaliștii se vor alătura din nou neuroștiințelor:

> Într-adevăr, biologia constituie domeniul unor posibilități nelimitate, de la care așteptăm explicațiile cele mai surprinzătoare; deocamdată nu putem ghici răspunsurile pe care le va da, după câteva decenii [...]. Ele vor fi, poate, de natură să năruie întreaga noastră construcție artificială de ipoteze[*3] [12].

Acesta nu era acel Freud extrem de speculativ despre care învățasem în facultate. „Proiectul” era o revelație pentru mine, cum fusese și pentru Freud. La vremea respectivă i-a scris prietenului său Wilhelm Fliess:

> Într-o noapte plină [...] barierele au fost brusc ridicate, vălurile au căzut și am putut să înțeleg de la detaliile nevrozelor la determinanții conștiinței. Totul părea să se potrivească, rotițele erau angrenate, obiectul îți dădea impresia că era într-adevăr o mașină care, curând, va merge de la sine[13].

Dar euforia a durat puțin. O lună mai târziu scria: „Nu mai înțeleg în ce stare eram când am conceput «Psihologia», nu reușesc să-mi dau seama cum am ajuns să v-o impun”[14]. Neavând metode neuroștiințifice adecvate, Freud s-a bazat pe „imagini, transpoziții și presupuneri” ca să-și traducă inferențele clinice mai întâi în termeni funcționali, apoi în termeni fiziologici și anatomici[15]. După o ultimă tentativă de revizuire (inclusă într-o lungă scrisoare trimisă lui Fliess pe 1 ianuarie 1896), „Proiectul” a dispărut din peisaj timp de vreo 50 de ani, după care a reapărut. Dar ideile pe care le conține („fantoma sa ascunsă”, cum spunea James Strachey, cel care i-a tradus opera în engleză) i-au influențat toate teoriile psihanalitice – în așteptarea progresului științific viitor[16].

„Proiectul” conținea două idei care ies în evidență în lumina descoperirilor contemporane. În primul rând, prozencefalul este un „ganglion simpatic” care monitorizează și reglementează nevoile corpului. În al doilea rând, aceste nevoi sunt forța motrice a vieții psihice, „motorul mecanismului psihic”[*4] [17]. În lipsa oricărei perspective neurobiologice asupra modului în care asemenea nevoi corporale sunt reglate în creier – ca să nu mai vorbim de modul în care ar putea fi explicate „cu ajutorul metodei fizico-matematice” –, Freud nu a putut decât „să presupună că există forțe noi, egale ca

importanță cu forțele chimico-fizice inerente materiei”, pentru a rămâne fidel idealurilor Societății de Fizică din Berlin. A numit „metapsihologice” aceste forțe *din spatele* fenomenelor psihologice. A precizat că dorea „să transforme metafizica în metapsihologie”[*5] [18]. Cu alte cuvinte, voia să înlocuiască filosofia cu știința – o știință a subiectivismului. Ne-a cerut să nu îi judecăm prea aspru inferențele speculative referitoare la procesele mintale latente:

> Această stare de lucruri se datorează numai faptului că am fost obligați să folosim anumiți termeni științifici, adică limbajului imagistic al psihologiei (mai exact, al psihologiei abisale). Fără ajutorul său nu am fi putut însă descrie procesele respective, mai mult, nici nu am fi putut să le identificăm. Probabil că minusurile descrierii noastre ar dispărea dacă am putea înlocui termenii psihologiei prin termeni din domeniul fiziologiei sau al chimiei[*6] [19].

Prima dintre noile forțe pe care Freud a fost obligat să le deducă era conceptul de „pulsiune”, pe care a definit-o drept „reprezentant psihic al stimulilor ce izvorăsc din interiorul corpului și ajung în suflet, ca *o măsură a cererii de travaliu ce e impusă psihicului* ca urmare a legăturii sale cu ceea ce e corporal”[*7] [20].

Noțiunea lui Freud de „pulsiune” – pe care o considera sursa întregii „energii psihice” – nu era diferită de „energia vitală” a lui Müller, dar se înrădăcina în nevoile corpului. Freud a descris mecanismele cauzale prin care pulsiunile devin cogniție intenționalã drept o „economie a forței nervoase”[21]. Totuși, a recunoscut singur că „aici s-a petrecut ceva complet insesizabil

pentru noi”[*8], cu referire la modul în care nevoile corpului devin energie psihică[22].

Când am citit aceste cuvinte, cam un secol mai târziu, am realizat că sosise vremea să înlocuim „termenii psihologiei cu termeni din domeniul fiziologiei sau al chimiei”. De pildă, forța pulsională din spatele viselor, care era „latentă” în relatările subiective ale pacienților lui Freud și a cărei existență era, de aceea, considerată nefalsificabilă, era în mod clar „manifestă” în dovezile obiective obținute prin metode fiziologice moderne in vivo care nu erau disponibile pe vremea lui Freud. Să luăm ca exemplu imaginile din figura 3, obținute prin tomografie cu emisie de pozitroni[23]. Ele arată clar că circuitul CĂUTĂRII „plin de aspirații” se aprinde ca un pom de Crăciun în timpul somnului cu vise, în timp ce lobii prefrontali inhibitori nu sunt activi. În baza acestor descoperiri, când am fost invitat împreună cu Hobson la conferința Science of Consciousness din 2006 să discutăm despre credibilitatea științifică a teoriei visului elaborate de Freud, colegii noștri au votat cu 2:1 în favoarea ideii că această teorie este din nou viabilă[24].

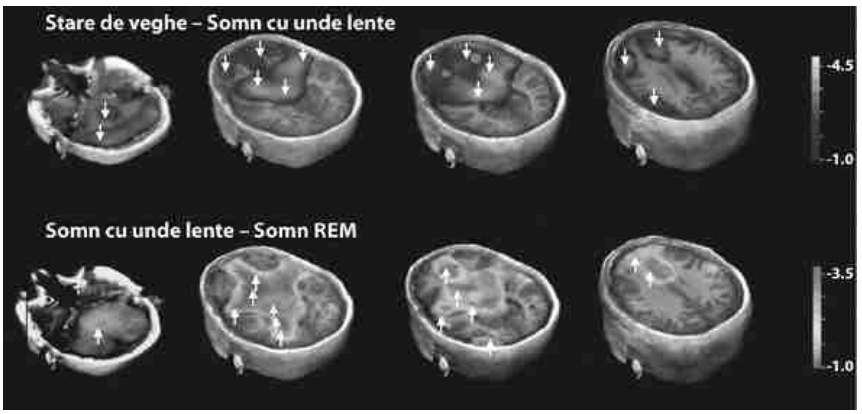


Figura 3. Rândurile orizontale arată secțiuni progresive în sus

ale creierului (de la stânga la dreapta). Rândul de sus arată diferența dintre starea de veghe și starea de somn, regiunea umbrită descriind activarea corticală *diminuată* odată cu debutul somnului; rândul de jos arată diferența dintre somnul REM și somnul non-REM (cu unde lente), regiunea evidențiată descriind activarea subcorticală *crescută* odată cu debutul somnului REM. Sistemul CĂUTĂRII se află în regiunea cu cea mai intensă activare.

„Eul” subiectiv nu a fost niciodată exclus din psihanaliză, în pofida neajunsurilor sale. Avea cea mai importantă poziție, oricât de stânjenitor era acest lucru pentru restul științei. Mulți dintre colegii mei din mediul academic m-au sfătuit să nu asociez ceea ce făceam cu psihanaliza, având în vedere bagajul istoric al acestui cuvânt. Mi-au spus că era ca și cum un astronom s-ar fi asociat cu astrologia. Dar eu am considerat că nu ar fi onest din punct de vedere intelectual dacă nu i-aș recunoaște meritele lui Freud. Indiferent cât de parţial şi-a atins obiectivele, ele erau în mod clar *obiectivele* corecte pentru o știință a minții. Așa că mi-am denumit abordarea „neuropsihanaliză”. Am spus că neuropsihologia care mi s-a predat la cursuri putea foarte bine să se fi numit neurobehaviorism, datorită atitudinii sale față de subiectivitate. Am vrut să fie clar că neuropsihologia pe care o elaboram se baza pe experiența trăită. În acest spirit, după ce am scris o lucrare programatică despre relația dintre psihanaliză și neuroștiințe, m-am pus pe treabă[25].

M-am mutat în Londra în 1989, ca să fac studii de

psihanaliză. Ca să-mi continui cercetările și activitatea clinică, am acceptat și poziția de *honorary lecturer* în neurochirurgie în cadrul Royal London Hospital School of Medicine. Am fost încântat să-i urmez puternica tradiție în neurologie: la mijlocul secolului al XIX-lea, John Hughlings Jackson, părintele fondator al neurologiei și neuropsihologiei britanice, profesa acolo ca medic. Spitalul Royal London se afla în Whitechapel, o zonă ce îi atrage de secole pe imigranți, deservind, astfel, dintotdeauna comunități vulnerabile. Îmi amintea de spitalul Baragwanath din Soweto. Mi-a fost ca o a doua casă.

La începutul anilor 1990, un coleg neurochirurg din Africa de Sud mi-a prezentat cazul domnului S., un pacient pe care, cu zece luni înainte, îl operase pentru a-i extirpa o tumoră ce îi creștea sub lobii frontali, deplasându-i nervii optici. Domnul S. suferise o mică hemoragie în timpul intervenției, care îi întrerupsese fluxul sangvin spre prozencefalul bazal (vezi figura 1). Nucleii prozencefalului bazal transmit acetilcolina către diversele structuri corticale și subcorticale implicate în accesarea memoriei pe termen lung. Se consideră că aceste căi colinergice interacționează cu căile dopaminergice (vezi figura 2), acestea din urmă fiind așa-numitul sistem de recompensă ce activează comportamentele de „căutare”, nu numai în legătură cu acțiunile fizice din lumea exterioară, ci și cu lumea interioară a reprezentărilor, acțiunile imaginare ce apar în gândire și în vise[26]. Din cauza acestei hemoragii, domnul S. s-a trezit din operație cu un sindrom amnezic profund, numit psihoză Korsakoff, a cărei caracteristică principală este o stare asemănătoare visului numită confabulație. Memoria sa legată de evenimentele recente era profund dereglată, astfel încât avea constant amintiri false. Acest deficit de *căutare* este destul

de debilitant, dar în cazul amneziei confabulatorii este agravat de faptul că pacienții nu își *monitorizează* adecvat veridicitatea amintirilor pe care le accesează în mod eronat, astfel tratându-le ca și cum ar fi reale, când, evident, nu sunt.

De exemplu, domnul S. credea că se află în Johannesburg (unde locuia), dar, de fapt, tocmai călătorise la Londra, ca să îl consult. Nu își amintea să fi călătorit. Când l-am corectat în această privință, a insistat că nu avea cum să fie în Londra. Așa că l-am rugat să se uite pe fereastră, întrucât ningea, iar asta nu se întâmplă niciodată în Johannesburg. Inițial a părut șocat, apoi s-a liniștit și a replicat: „Nu, *știu* că sunt în Johannesburg – dacă mănânci pizza nu înseamnă că și ești în Italia".

Domnul S. era inginer electronist şi avea 56 de ani. L-am consultat zilnic la clinica mea pentru tratament ambulatoriu de zi, de șase ori pe săptămână, în încercarea de a-l orienta și a-l ajuta să înțeleagă modul în care îl trăda memoria. Deși îl consultam zilnic în același loc și la aceeași oră, nu m-a recunoscut niciodată drept medicul lui de la o întâlnire la alta. Aparent, îmi recunoștea chipul, dar mă confunda de multe ori cu persoane pe care le știa din alte contexte – de obicei cu un coleg inginer care lucra cu el la vreo problemă de electronică sau cu un client care avea nevoie de ajutorul lui de specialitate. Cu alte cuvinte, domnul S. mă trata ca și cum eu aș fi avut nevoie de ajutorul lui, nu invers. De asemenea, își imagina frecvent că eram amândoi studenți și că ieșiserăm la un pahar după vreo activitate sportivă (un concurs de canotaj sau un meci de rugbi). La vremea respectivă eu eram suficient de tânăr ca așa ceva să fie plauzibil, dar domnul S. terminase facultatea de mai bine de 30 de ani.

După fiecare consult clinic, discutam separat cu soția lui ca

să contextualizez amintirile lui eronate și să încerc să le găsesc semnificația. Aceasta era principala diferență dintre abordarea mea și cea a colegilor mei, mai tradițională, față de „reabilitarea cognitivă”. Neuropsihologii sunt preocupați de obicei de *gradul* tulburării de memorie, evaluat din perspectiva persoanei a treia, însă pe mine mă interesa mai mult *conținutul* subiectiv al erorilor domnului S., înțeles din perspectiva persoanei întâi. Am plecat de la ipoteza că semnificația personală a evenimentelor ce îi veneau obsesiv în minte, în locul amintirilor-țintă pe care le căuta, avea să clarifice cât de cât mecanismul acestor confabulații, deschizând astfel căi noi spre influențarea lor. Așadar, la întâlnirile mele cu soția lui, de exemplu, voiam să știu dacă domnul S. chiar făcuse parte din vreo echipă de canotaj și de rugbi în studenție și dacă oferea într-adevăr asistență de specialitate în probleme de electronică.

Două lucruri pe care le-am aflat astfel sunt relevante pentru înțelegerea confabulațiilor sale. Mai întâi, e vorba de faptul că suferise cândva de nişte probleme cronice la dantură – probleme tratate (cu succes) în cele din urmă, cu ajutorul unor implanturi dentare – și, în al doilea rând, că suferea de aritmie cardiacă, ţinută sub control cu ajutorul unui stimulator cardiac.

Am selectat un scurt fragment din transcrierea unei înregistrări audio a primelor câteva minute din a zecea consultație pe care am avut-o cu domnul S. Am ales acest fragment deoarece, când m-am dus să-l preiau din sala de așteptare în ziua aceea, a părut pentru scurt timp (și pentru prima dată) că recunoștea cine sunt și de ce venise la mine la consultație. Când am intrat în sala de așteptare, și-a atins cicatricea de la craniotomie din vârful capului și a spus: „Bună, doctore”.

Ne-am instalat în cabinetul meu, eu sperând să aflu mai multe, profitând de acest licăr de introspecție, dacă despre așa ceva era vorba.

Eu: Ați dus mâna la cap când ne-am văzut în sala de așteptare.
Domnul S.: Cred că problema e că lipsește un cartuș. Trebuie să… ne trebuie doar specificațiile. Ce era? Un C49? Ar trebui să-l comandăm?
Eu: Ce face un cartuș C49?
Domnul S.: Memorie. E un cartuș de memorie, un implant de memorie. Dar n-am înțeles niciodată cum funcționează. De fapt, nu l-am mai folosit de vreo cinci-șase luni. Se pare că n-avem nevoie de el de fapt. A fost scos tot de un doctor. Cum îl cheamă? Cred că Solms. Dar se pare că nu-mi trebuie de fapt. Implanturile funcționează bine.
Eu: Vă dați seama că ceva nu e în regulă cu memoria dumneavoastră, dar…
Domnul S.: Da, nu funcționează 100%, dar nu ne trebuie de fapt – doar avea o mică defecțiune, îi lipseau niște părți. Analiza arată că lipsea un C sau un C09. Denise [prima lui soție] m-a adus aici să mă vadă un doctor. Cum îl cheamă? Iar am uitat. Solms sau așa ceva. Și a făcut un transplant din ăla de inimă și acum merge bine, nu ratează nicio bătaie.
Eu: Vă dați seama că ceva lipsește. Unele amintiri lipsesc și, desigur, asta vă îngrijorează. Sperați că eu pot să rezolv problema, așa cum ceilalți doctori v-au rezolvat problemele cu dantura și cu inima. Dar vă doriți acest lucru atât de mult, încât vă e greu să acceptați faptul că încă nu s-a rezolvat.
Domnul S.: A, înțeleg. Da, nu merge 100%. [Își atinge capul.]

> M-am lovit la *c*ap. Am ieșit de pe teren câteva minute. Dar acum e bine. Presupun că n-ar trebui să mă întorc. Dar mă știi doar, nu-mi place să stau pe tușă. Așa că l-am întrebat pe Tim Noakes [celebru medic sportiv din Africa de Sud] – că asigurarea medicală o am, știi, așa că de ce să n-o folosesc, de ce să nu mă duc la cel mai bun – și el a spus: „Bine, joacă mai departe”.

Mă opresc aici. Ar trebui să fie suficient de ușor de recunoscut tulburările pur cognitive ale căutării și monitorizării memoriei pe care le-am menționat mai sus. Când domnul S. m-a văzut intrând în sala de așteptare pentru cea de-a zecea consultație, apariția mea i-a evocat o mulțime de asocieri – legate de doctori, capul său, amintirile lipsă, intervențiile chirurgicale și altele asemenea. Dar în niciunul dintre aceste cazuri nu a accesat exact amintirea-țintă pe care o căuta. În schimb, au apărut cele ce pot fi caracterizate drept „ţinte potenţiale” – amintiri din aceleași categorii semantice largi ca și țintele, dar localizate greșit în timp și spațiu. Astfel, ideea de „doctor” a evocat asocieri referitoare la un neurochirurg și la un celebru medic sportiv în locul țintei, eu; ideea de „cap” a evocat o situație când s-a lovit la cap în locul unei tumori cerebrale; „amintirea lipsă” a evocat un cartuș electronic în locul amneziei lui; „intervențiile chirurgicale” i-au evocat intervențiile dentare și cardiologice anterioare în locul intervenției chirurgicale recente pe creier și așa mai departe. Deficitul de monitorizare e la fel de simplu de observat: domnul S. a acceptat veridicitatea amintirilor sale eronate cu prea multă uşurinţă. Faptul că se simţea ca un student de douăzeci și ceva de ani pe un teren de rugbi (în pofida tuturor dovezilor

contrare) este un exemplu evident în acest sens. La fel și convingerea lui că încă se afla în Johannesburg.

Dar, când analizăm confabulațiile domnului S. dintr-un punct de vedere *subiectiv*, apar și alte fapte. Imaginează-ți cum te-ai *simți* dacă ai realiza brusc că nu îl recunoști pe medicul care tocmai a intrat în încăpere, deși el pare să fie responsabilul pentru serviciile medicale pe care le primești, că nu știi în ce încăpere (sau chiar în ce oraș) te afli, că ai o cicatrice mare în vârful capului și nu știi cum a apărut acolo, că – de fapt – nu îți amintești ce s-a întâmplat cu doar două minute în urmă, ca să nu mai vorbim de zilele și lunile dinaintea momentului prezent. Probabil ai simți un soi de *panică*, întrebându-te dacă nu cumva acest medic ți-a făcut o operație la cap, în urma căreia nu îți mai amintești nimic de la un moment la altul. Iată cum resimte subiectul intențional al minții – eul viu – lipsa mecanismelor de căutare și de monitorizare a amintirilor.

Să observăm acum ce a făcut domnul S. ca urmare a acestor sentimente (cu alte cuvinte, să observăm efectele lor de cauzalitate asupra cogniției lui). Când a realizat că îi lipsește „cartușul de memorie”, s-a liniștit (printr-un delir) că *poți pur și simplu să-ți comanzi unul nou*. Pentru că nu a fost cu totul convins de propriile argumente, s-a răzgândit. De fapt, *nu ai neapărat nevoie de cartuș*, poți să te descurci și fără, iar el deja se descurcă fără probleme de luni de zile. Apoi a făcut o legătură între cartușul lipsă și cicatricea de la craniotomie: se pare că un doctor i-a tăiat ceva. Speră că nu e vorba de doctorul care stă în fața lui și, mai mult, speră că intervenția nu a fost un eșec. În acest moment, domnul S. își amintește că intervenția chirurgicală dentară și cea cardiacă prin care a trecut au avut succes și contopește (printr-un delir) aceste intervenții cu cea

de la cap: *a fost reușită, implanturile funcționează bine* și „nu ratează nicio bătaie”. Când introduc o sămânţă de îndoială în partitura lui, schimbă direcţia. E de acord că nu funcționează 100%, dar, în același timp, hotărăște că ceea ce i s-a întâmplat la cap nu a fost intervenție chirurgicală până la urmă, *ci doar s-a lovit* – acum este afectat de efectele temporare ale unei accidentări minore pe terenul de sport. În consecință, a fost trimis pe tușă câteva minute. Dar, din fericire, pentru că își permite să achite o consultație la cel mai bun medic sportiv, se liniștește din nou: *poate juca mai departe*. Totul va fi bine.

Analizarea confabulațiilor domnului S. din perspectiva persoanei întâi ne relevă în mod clar ceva nou despre ele: conținutul amintirilor sale eronate are un caracter puternic *motivat*. Ele sunt departe de a fi nişte erori întâmplătoare de căutare, conțin un bias cognitiv clar – au obiectivul și scopul de a-i reîncadra situația încărcată de anxietate drept una liniștitoare, sigură și familiară. Așadar, după cum a dedus și Freud în cazul viselor, confabulațiile sunt motivate. Procesele psihice din amnezia confabulatorie vin din niște *aspirații*. Dar acest fapt devine evident numai atunci când contextul emoțional și semnificația personală (trăită numai de domnul S.) ale implanturilor dentare („implanturile funcționează bine”) și stimulatoarelor cardiace („nu ratează nicio bătaie”) sunt luate în considerare – cum ar proceda un psihanalist. Iată ce nu reușesc să vadă neuropsihologii când urmăresc să fie cu totul obiectivi; cum spunea Sacks, când exclud psihicul.

Perspectiva observației la persoana întâi pe care tocmai am descris-o relevă și ceva nou despre *mecanismul* confabulației, ceva trecut cu vederea din perspectiva persoanei a treia. Ea ne spune că confabulația nu apare exclusiv din cauza deficitelor

de căutare strategică și de monitorizare a surselor („cartușele de memorie” lipsă), ci și din cauza dezinhibării unor forme de reamintire mediate într-un mod mai emoțional, cam așa cum funcționează memoria unui copil. Acest mecanism psihodinamic are implicații pentru tratamentul confabulației și, desigur, pentru întrebarea referitoare la procesele cerebrale pe care le implică ea. Se dovedește că funcțiile corecte de căutare și de monitorizare ale memoriei depind parțial de circuitele colinergice din prozencefalul bazal, care constrâng mecanismele de „recompensă” ale circuitului dopaminergic mezocortical-mezolimbic în accesarea memoriei. Se întâmplă că în cazul viselor apare o eliberare similară a căutării dopaminergice[27]. Din acest motiv, le-am prezentat colegilor mei cazul domnului S. sub titlul „Omul care trăia într-un vis”.

Acest lucru mi-a permis, ca în cazul viselor, să leg provizoriu mecanismul dopaminergic neîngrădit al „recompensei”, al „dorinței” sau al „CĂUTĂRII” de noțiunea lui Freud de „împlinire a dorințelor”[28] – un concept metapsihologic strâns legat de conceptul său de „pulsiune”[29]. Dimpotrivă, funcțiile nucleilor colinergici din prozencefal pot fi legate din anumite puncte de vedere de influențele inhibitorii ale „testării realității”[30]. În acest fel, am început să traduc deducțiile lui Freud despre mecanismele funcționale ale subiectivității prin echivalentele lor fiziologice.

Aceștia au fost primii mei pași. În mod firesc, asemenea generalizări ample nu se pot baza pe dovezi pur clinice dintr-un singur caz. Prin urmare, după ce mi-am formulat impresia despre domnul S., am cooptat evaluatori „în orb” (colegi nefamiliarizați cu ipoteza mea) pentru a măsura, pe o scală Likert în șapte puncte, gradul de agreabilitate versus

dezagreabilitate într-un eșantion continuu neselectat alcătuit din 155 dintre confabulațiile pacientului. Rezultatele au fost (foarte) semnificative statistic: când erau comparate cu amintirile-țintă pe care le-au înlocuit, confabulațiile domnului S. îi îmbunătățeau substanțial situația din punct de vedere emoțional[31]. În continuare, am demonstrat împreună cu colaboratorii mei același efect puternic în studiile în care erau implicați numeroși alți pacienți cu confabulații. Studiile empirice ulterioare au validat statistic efectele confabulației de reglare a dispoziției afective pe care le-am dedus clinic în cazul domnului S.[32]. Acest program de cercetare a inițiat o abordare cu totul nouă a neuropsihologiei confabulației[33] și a tulburărilor conexe precum anosognozia[34]. De asemenea, a pus bazele unei abordări noi a tulburărilor psihiatrice comune precum dependența și depresia majoră[35]. De trei decenii lucrez la dezvoltarea acestei abordări „neuropsihanalitice” a bolilor mintale, încercând să readuc subiectivitatea în neuroștiințe[36].

Pe măsură ce am acumulat experiență clinică de tipul celei tocmai descrise în timpul specializării mele în psihanaliză la Londra, am fost invitat să-mi expun constatările într-o serie de prezentări științifice la New York. Acestea au început în 1992 cu un simpozion de o zi la Academia de Medicină din New York și au continuat sub forma unor seminare lunare cu cei mai apropiați colaboratori ai mei la Societatea și Institutul de Psihanaliză din New York[37]. Treptat, întâlnirile acestui grup de colegi au generat activități similare şi în alte părți ale lumii, ducând la decizia (în 1999) de a înființa o revistă nouă care să servească drept vehicul pentru comunicarea dintre noi. Revista avea nevoie de un nume și așa am folosit pentru prima dată

termenul inventat de mine, *Neuropsychoanalysis*.

Activitatea mea în acest domeniu interdisciplinar a primit încurajări puternice din partea lui Eric Kandel, pe care l-am cunoscut în 1993. Kandel este atipic printre specialiștii în neuroștiințe prin marele său respect pentru Freud. De fapt, inițial intenționase să se specializeze în psihanaliză, dar a fost descurajat de Ernst Kris, un analist important la momentul respectiv și tatăl iubitei de atunci a lui Kandel. Acesta din urmă nu ascunde faptul că Kris considera că nu i se potrivea practica clinică în psihiatrie și cred că îi este recunoscător că l-a îndrumat spre cercetare în domeniu.

La cinci ani după ce l-am cunoscut (și cu doi ani înainte să i se acorde Premiul Nobel), Kandel a publicat un articol intitulat „A new intellectual framework for psychiatry", în care afirma că psihiatria secolului XXI ar trebui să se bazeze pe integrarea neuroștiințelor și psihanalizei[38]. Într-un articol ulterior, a afirmat: „Psihanaliza încă reprezintă viziunea asupra minții cea mai coerentă și mai satisfăcătoare din punct de vedere intelectual"[39]. Această opinie coincidea cu a mea: în pofida tuturor neajunsurilor sale, psihanaliza constituie în prezent cel mai bun punct de plecare sub aspect conceptual pentru abordarea științifică a subiectivității.

Așadar, nu m-a surprins când Kandel mi-a acceptat invitația de a face parte din consiliul editorial fondator al revistei *Neuropsychoanalysis*, alături de o masă critică de alți specialiști în neuroștiințe și psihanaliști importanți, care au fost de acord că acesta era drumul corect pe care trebuiau să-l urmeze disciplinele noastre[40]. Un an mai târziu, la începutul noului secol, am fondat Societatea Internațională de Neuropsihanaliză, cu Jaak Panksepp și cu mine ca primi copreședinți. Acest

eveniment a coincis cu congresul inaugural al societății, care, de atunci, se ține anual în diferite orașe din lume. Tema primului congres au constituit-o emoțiile. Întâlnirea a avut loc la Royal College of Surgeons of England, invitații fiind Oliver Sacks, Jaak Panksepp, Antonio Damasio și eu.

Am menționat deja relația mea cu Oliver Sacks. Am vorbit și despre lucrarea lui Jaak Panksepp *Affective Neuroscience*, al cărei titlu face aluzie la faptul că neuroștiința cognitivă nu acorda atenție suficientă „afectului" (termenul tehnic pentru sentimente). După ce am citit-o, am stabilit o colaborare științifică strânsă cu Panksepp, ceea ce a deplasat decisiv centrul atenției în activitatea mea în următoarele două decenii de la cortex la trunchiul cerebral. Îi sunt profund îndatorat pentru că mi-a arătat calea spre ideile despre care voi vorbi în paginile următoare, acesta fiind motivul pentru care îi dedic cartea de față.

M-am familiarizat mai întâi cu lucrările lui Antonio Damasio și ale partenerei sale, Hanna Damasio, în timpul studiilor mele de neuropsihologie. Erau nişte specialiști în neuroștiințele cognitive foarte respectați, al căror manual, *Lesion Analysis in Neuropsychology* (1989), a fost indispensabil pentru cercetările mele asupra viselor. Totuși, ceea ce l-a făcut pe Damasio cunoscut în toată lumea a fost volumul său *Eroarea lui Descartes* (1994), o pledoarie înflăcărată adresată neuroștiinței cognitive pentru o mai mare valorizare a afectului.

De asemenea, Panksepp și Damasio au jucat un rol important în cadrul celui de-al Doisprezecelea Congres Internațional de Neuropsihanaliză din 2011, care s-a dovedit a fi un punct de cotitură în domeniu. Congresul s-a desfășurat la Berlin, pe tema creierului in vivo. Ceilalți invitați au fost Bud

Craig și Vittorio Gallese – doi dintre principalii experți din lume pe acest subiect. De obicei, rolul meu la aceste congrese este să prezint observaţiile finale, ce rezumă temele principale și, mai important, *integrează* perspectivele din neuroștiințe și psihanaliză prezentate. De data aceasta, sarcina mea a fost neobișnuit de dificilă.

Prima dificultate venea din faptul că Damasio se contrazisese puternic cu Craig în timpul congresului cu privire la modul în care apare în creier „sinele" care simte. Deși amândoi erau de acord că simţul sinelui apare din regiunile cerebrale care monitorizează starea corpului visceral, Damasio – în spiritul lui Panksepp – considera că mecanismele în cauză se găsesc cel puțin parțial în trunchiul cerebral. Craig, dimpotrivă, pretindea că se localizează exclusiv în cortex – în insula anterioară, mai exact. Acest dezacord a fost relativ simplu de rezolvat în concluziile mele de final, deoarece Damasio furnizase date convingătoare, axându-se pe un pacient căruia i se îndepărtase complet cortexul insular. Voi descrie acest pacient în capitolul următor.

Mult mai dificil de reconciliat a fost o contradicție majoră apărută în timpul congresului între perspectivele noi ale lui Panksepp și Damasio, pe de o parte, și cele vechi ale lui Freud, pe de altă parte.

Pacientul lui Damasio fără cortexul insular „raporta senzații de foame, sete și dorința de a excreta și se comporta în consecință"[41]. Aceste senzații sunt exemple a ceea ce Panksepp numește „afecte homeostatice" – afecte ce reglează nevoile vitale ale corpului. Freud le numea „pulsiuni" – sursa „energiei psihice", „motorul mecanismului psihic". Termenul general al

lui Freud pentru partea minții care îndeplinește aceste funcții vitale era „Se”:

> Se-ul, separat de lumea exterioară, își are propria sa lume perceptivă. El simte cu o acuitate extraordinară anumite transformări din interiorul său, mai cu seamă oscilațiile din tensiunea nevoilor a pulsiunilor lui, care devin conștiente ca senzații ale seriei plăcere-neplăcere. Este, ce-i drept, greu de indicat pe ce căi și cu ajutorul căror organe terminale sensibile iau naștere aceste percepții. Dar este cert că autopercepțiile – afecte generale și senzații de plăcere-neplăcere – domină procesele din Se cu o forță despotică. Se-ul ascultă de principiul neînduplecat al plăcerii[*9] [42].

După cum am spus, obiectivul meu științific era să traduc asemenea noțiuni metapsihologice în limbajul anatomiei și al fiziologiei, astfel încât să pot integra în neuroștiințe abordarea lui Freud. Dar aici m-am poticnit într-o contradicție majoră a concepției clasice a lui Freud: el ajunsese la concluzia că „Se-ul” este inconștient. Aceasta a fost una dintre concepțiile lui fundamentale cu privire la modul în care funcționează mintea. Îmi era clar că partea creierului ce măsoară „cererea de travaliu ce e impusă psihicului ca urmare a legăturii sale cu ceea ce e corporal” – partea care generează ceea ce Freud numea „pulsiuni”, care sunt sinonime cu „afectele homeostatice” ale lui Panksepp (ce declanșează mecanismul lui de CĂUTARE venind din aspirații) – se afla în trunchiul cerebral și în hipotalamus (vezi figura 1). Aceasta este partea creierului care se supune „principiului plăcerii”. Dar cum pot fi inconștiente senzațiile de plăcere? După cum am văzut la

pacientul lui Damasio, pulsiunile precum foamea, setea și dorința de a excreta sunt *simțite.* Desigur că așa e. Totuși, Freud a spus că Se-ul – sediul pulsiunilor – este inconștient. Asimilase aceeași doctrină clasică precum Craig (ca și mine, cel puțin inițial) și, prin urmare, localizase conștiința în cortexul cerebral. Așa că a scris următoarele în eseul său din 1920 pe care l-am citat mai sus şi în care spera că neajunsurile din teoriile sale vor dispărea odată ce vom putea înlocui termenii psihologici cu termeni din fiziologie și din chimie:

> Cum conștiința realizează în special percepții ale unor excitații care provin din lumea exterioară și senzații de plăcere sau neplăcere, care nu-și pot avea sursa decât în interiorul aparatului psihic, este posibilă o situare spațială a sistemului *P-Cs* (percepție-conștiință). El trebuie să se afle la granița care separă lumea exterioară de cea interioară, să fie orientat către lumea exterioară și să înglobeze toate celelalte sisteme psihice. Observăm însă că, făcând aceste presupuneri, nu am spus nimic nou, ci *ne-am raliat anatomiei localizante a creierului care situează „sediul" conștiinței în scoarța cerebrală*, stratul cel mai exterior, periferic al organului central. Anatomia creierului nu are motive să se întrebe de ce, anatomic vorbind, conștiința este localizată tocmai la suprafața creierului, și nu într-o regiune mai protejată, undeva în straturile intime ale acestuia[*10] [43].

În caz că mai există vreo îndoială în legătură cu opinia lui Freud că *toată* conștiința (inclusiv senzațiile de plăcere și neplăcere) este localizată în cortex, mai dau un citat:

> Conștientizarea este legată mai cu seamă de percepțiile pe

care organele noastre senzoriale le capătă de la lumea exterioară. Este deci pentru considerația topică un *fenomen care se desfășoară în stratul cel mai exterior al scoarței Eului*. Noi primim, desigur, și știri conștiente din interiorul corpului, sentimentele, care chiar influențează substanțial viața noastră psihică decât percepțiile exterioare și, în anumite împrejurări, transmit și organelor senzoriale sentimente, senzații de durere, în afară de percepțiile lor specifice. Dar, întrucât aceste senzații, cum se numesc ele spre deosebire de percepțiile conștiente, emană tot de la organele terminale și *noi le concepem pe toate ca prelungiri, ca derivate ale stratului cortical*, putem menține în picioare afirmația de mai sus. Deosebirea ar fi numai că, în ceea ce privește organele terminale ale senzațiilor și sentimentelor, însuși corpul ar înlocui lumea exterioară[*11] [44].

În mod clar, pentru Freud, sentimentele conștiente, nu mai puțin decât percepțiile, sunt generate în „Eu” (acea parte a minții pe care a identificat-o cu cortexul)[45], și nu în „Se-ul” inconștient – pe care eram acum obligat să-l localizez în trunchiul cerebral și hipotalamus. Pe scurt, părea că Freud a înțeles invers relația funcțională dintre „Se” (trunchiul cerebral) și „Eu” (cortex), cel puțin în ce privește sentimentele. El credea că Eul care percepe este conștient, iar Se-ul care simte este inconștient. Se poate oare să fi înțeles el pe dos modelul minții[46]?

[*1] Sigmund Freud, *Opere*, vol. XVII: *Proiect de psihologie*, traducere din limba germană de Daniela Ştefănescu, cuvânt înainte de Vasile Dem. Zamfirescu. București: Trei, 2007, p.

31 (n.tr.).

*2 Sigmund Freud, „Pentru a introduce narcisismul”, traducere de Gilbert Lepădatu, în *Opere esenţiale*, vol. 3: *Psihologia inconştientului*, notă asupra ediţiei de Raluca Hurduc, note introductive de Roxana Melnicu. Bucureşti: Trei, 2017, e-book (n.tr.).

*3 Sigmund Freud, „Dincolo de principiul plăcerii”, în *Opere esențiale*, vol. 3: *Psihologia inconștientului*, traducere de G. Purdea și Vasile Dem. Zamfirescu. București: Trei, e-book (n.tr.).

*4 Sigmund Freud, *Proiect de psihologie*, ed.cit., pp. 41, 55 (n.tr.).

*5 Sigmund Freud, *Psihopatologia vieţii cotidiene. Despre uitare, greşeli de vorbire, superstiţie şi eroare*, traducere de Herta Spuhn şi Daniela Ştefănescu, ediţia a II-a. Bucureşti: Trei, 2010, e-book (n.tr.).

*6 Sigmund Freud, „Dincolo de principiul plăcerii”, ed.cit. (n.tr.).

*7 Sigmund Freud, „Pulsiuni și destine ale pulsiunilor”, în *Opere esenţiale*, vol. 3: *Psihologia inconştientului*, ed.cit. (n.tr.).

*8 Sigmund Freud, „Compendiu de psihanaliză”, în *Despre psihanaliză*, traducere din germană de Daniela Ştefănescu, prefaţă de Vasile Dem Zamfirescu. Bucureşti: Trei, 2014, p. 241 (n.tr.).

*9 *Ibidem*, p. 243 (n.tr.).

*10 Sigmund Freud, „Dincolo de principiul plăcerii”, ed.cit. (n.tr.).

*11 *Ibidem*, p. 193 (n.tr.).

3

Eroarea privitoare la cortex

Pe la sfârșitul anului 2004, Björn Merker, un specialist în neuroștiințe, a plecat împreună cu cinci familii cu copii cu afecțiuni neurologice într-o excursie de o săptămână la Disney World. Copiii aveau vârste între zece luni și cinci ani. Au ales diverse atracții: una dintre preferate a fost „It's a Small World After All". Au făcut poze cu Mickey Mouse. Au mâncat popcorn, crenvurști și înghețată – poate mai mult decât era cazul – și au băut mult suc. Uneori, păreau că sunt copleșiți, au plâns de multe ori. Dar, în pofida acestor momente stresante, doctorul Merker a fost frapat de cât de mult păreau să se distreze copiii – cât de mult le plăcea că sunt acolo. La urma urmei, conform uneia dintre premisele fundamentale ale neurologiei, acești copii ar fi trebuit să fie într-o „stare vegetativă". Ar fi trebuit să fie la un pas de comă: capabili să manifeste doar funcții autonome precum reglarea bătăilor inimii, respirația și activitatea gastrointestinală, răspunsurile lor motoare limitându-se la simple reflexe precum clipitul sau înghițitul.

Merker se împrietenise cu aceste familii în anul precedent, când s-a alăturat unui grup de întrajutorare pentru îngrijitorii copiilor care suferă de o afecțiune cerebrală rară numită

hidroanencefalie. De pe la începutul lui 2003, Merker citise peste 26.000 de e-mailuri schimbate între membrii grupului. Având argumente teoretice, se îngrijorase că exista posibilitatea ca presupunerea generală că acești copii vor fi „vegetativi” să fi devenit o profeţie care se autoîmplineşte nu din cauza afecțiunii în sine, ci a faptului că asemenea pacienți erau *tratați* de majoritatea pediatrilor și neurologilor ca și cum ar fi fost complet insensibili. Iată ce a raportat:

> Acești copii sunt nu numai treji și adesea vioi, ci manifestă şi receptivitate la mediul ambiant sub forma unor reacții emoționale sau de orientare la evenimente din mediu [...]. Mai mult, copiii au epilepsie cu crize de absență. Părinții recunosc aceste întreruperi de accesibilitate la copiii lor, comentând despre ele în termeni precum „ia o pauză să vorbească cu îngerii” și dându-şi seama foarte uşor când copilul „și-a revenit”. [...] Faptul că acești copii prezintă asemenea episoade pare să fie o dovadă serioasă a stării lor de conștiință[1].

Cea mai importantă observație a lui Merker nu este că pacienții își pierd și își recapătă luciditatea, ci că manifestă „receptivitate la mediul ambiant sub forma unor reacții emoționale sau de orientare la evenimente din mediu”. Aceasta este trăsătura definitorie a ceea ce se presupune că le lipsește pacienților vegetativi: intenționalitatea. Din acest motiv, starea vegetativă este definită și ca „veghe *fără reacţie*”[2]. În general, se consideră că acești pacienți înregistrează stimuli vizuali, auditivi și tactili *în mod inconștient*, dar Merker i-a văzut pe copii exprimându-şi plăcerea prin zâmbete și râsete, iar

aversiunea prin proteste, arcuiri ale spatelui și plânsete, „chipurile fiindu-le animate de aceste stări emoționale”. I-a observat pe adulții familiari copiilor folosindu-se de reacțiile lor pentru a construi secvențe de joacă, copiii progresând previzibil de la zâmbete la chicoteli, râsete și mult entuziasm. Copiii răspundeau cel mai puternic la vocile și acțiunile părinților și ale altor persoane cu care erau familiarizați și manifestau preferințe pentru anumite situații în detrimentul altora. De exemplu, păreau să se bucure de anumite jucării, cântece sau videoclipuri și chiar ajungeau să aștepte prezența regulată a unor asemenea lucruri în programul lor zilnic. Deși comportamentul varia de la un copil la altul, unii dintre ei manifestau în mod clar inițiativă (în limitele dizabilităților lor motorii) – de exemplu, lovind obiecte care făceau zgomot agățate într-un cadru special construit în acest scop sau activând jucăriile preferate utilizând un întrerupător. Astfel de comportamente erau însoțite de semne de plăcere sau de entuziasm adecvate situației.

E clar că acești copii nu pot fi descriși drept vegetativi. Toate acestea sunt cu atât mai uimitoare cu cât copiii cu hidroanencefalie *se nasc fără cortex*. De obicei, cauza este un accident vascular masiv în uter, ce are drept consecință reabsorbția prozencefalului, astfel încât craniul fătului se umple cu lichid cefalorahidian în loc de țesut cerebral. De aici și termenul „hidroanencefalie”, care înseamnă „apă în loc de encefal”. O ilustrare a acestei afecțiuni găsiți în figura 4: un RMN cerebral al unei fetițe de trei ani născute fără cortex cerebral. Fotografiile înfățișând un asemenea copil (figura 5) demonstrează reacţia ei emoțională când frățiorul său îi este așezat în poală.

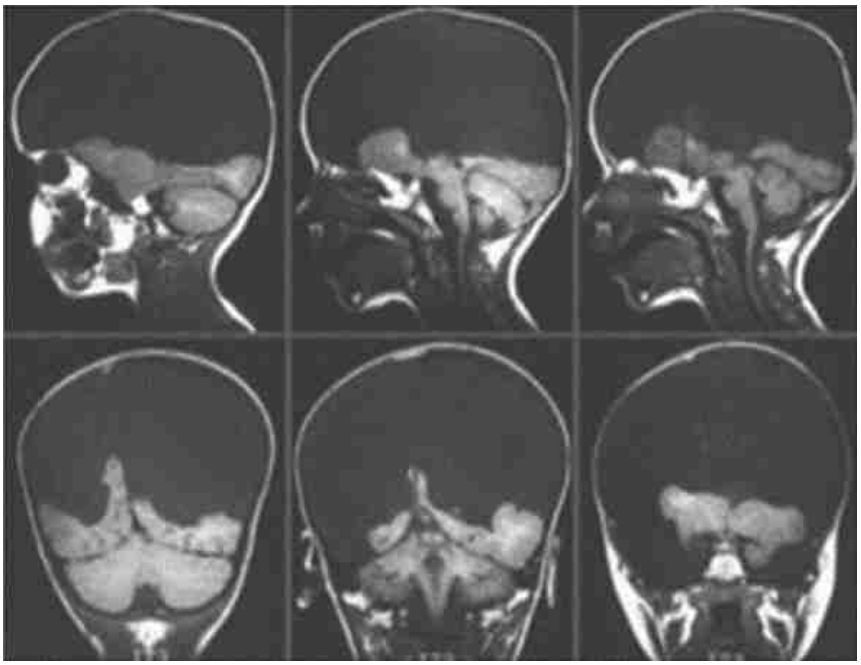


Figura 4. RMN cerebral al unei fetițe de trei ani născute fără cortex cerebral. Regiunea întunecată extinsă din interiorul craniului indică lipsa țesutului.

Figura 5. Reacția unei fetițe cu hidroanencefalie când frățiorul ei îi este așezat în poală

Unde este generată conștiința în creier? De 150 de ani, răspunsul aproape universal la această întrebare este „în cortex”. Aceasta a fost singura idee asupra căreia Freud și curentul tradițional principal din știința minții din secolul XX au căzut de acord. Și totuși, dacă această afirmație este valabilă, atunci, când cortexul lipsește, conștiința ar trebui să dispară. În cazul copiilor cu hidroanencefalie, acest lucru nu pare să se întâmple. Toate dovezile comportamentale sugerează că ei sunt de fapt conștienți. Nu sunt comatoși și nici nu trăiesc într-o stare vegetativă.

Oare trebuie respinsă teoria conform căreia conștiința ar fi localizată în cortex? Propun să nu ne grăbim să tragem această concluzie. O obiecție posibilă este că, la acești copii, cortexul nu a fost *îndepărtat* chirurgical la propriu. Procedura respectivă se numește „decorticare”.

Ce se întâmplă în cazul acestei intervenții? Evident, asemenea proceduri experimentale nu pot fi efectuate pe copii, dar s-au realizat adesea la nou-născuți ai altor mamifere, precum câini, pisici și șobolani. Rezultatul este întotdeauna același: conform criteriilor comportamentale obiective pe care le folosim în mod obișnuit ca să o măsurăm, conștiința *se menține*[*1]. Comportamentul postoperator al acestor animale nu poate fi descris drept „vegetativ” sau „comatos” oricât am lărgi definițiile acestor termeni. Merker scrie că acestea nu prezintă „anomalii comportamentale majore ce i-ar permite unui observator obișnuit să le identifice ca având dizabilități”. Antonio Damasio e de acord: „Mamiferele decorticate prezintă o persistență remarcabilă a comportamentului coerent, orientat spre obiective corespunzător senzațiilor și conștiinței”[3]. De pildă, șobolanii supuși procedurii de decorticare după naștere

stau pe patru membre, se ridică, se caţără, se agaţă de bare și dorm în poziții normale. Se curăţă, se joacă, înoată, mănâncă și se apără. Indiferent de sex, asemenea șobolani sunt capabili să se împerecheze cu succes cu parteneri normali ținuți în cușcă. Când cresc, femelele manifestă elementele esențiale ale comportamentului matern, care, deși deficitar în anumite privințe, le permite să-și crească puii până la maturitate[4].

Situația este și mai ciudată decât pare inițial. Din multe puncte de vedere, mamiferele supuse decorticării cerebrale sunt de fapt *mai* active, *mai* emoționale și *mai* receptive decât cele normale. Panksepp obișnuia să le ceară studenților să aleagă între două grupuri de șobolani, determinând în funcție de comportament grupul ai cărui membri fuseseră supuși intervenției. De obicei, studenții îi alegeau pe cei normali, gândind că șobolanii din celălalt grup (în cazul căruia se făcuse decorticarea) erau „mai vioi”[5].

Dacă susținem ipoteza conform căreia conștiința ar fi localizată în cortex, atunci aceste animale vioaie – și copiii expresivi, cu reacții emoționale observați de Merker la Disney World – trebuie să fie într-un anumit sens inconștiente. Cum ar fi posibil acest lucru?

Există un răspuns convențional la această întrebare. Iată povestea. Într-un anumit sens, există doi creieri, care îndeplinesc aceleași funcții și care comunică între ei în anumite momente, dar care, altfel, nu au deloc același statut. Unul (cortexul) este psihologic și conștient. Celălalt (trunchiul cerebral) nu este deloc aşa. Informațiile de la organele de simț sunt transmise nu numai cortexului, ci și coliculilor superiori ai trunchiului cerebral, printr-un set de conexiuni subcorticale (vezi figura 6). Aceste conexiuni prelucrează informațiile

senzoriale, dar o fac în mod *inconștient*. Să luăm exemplul celebrului fenomen al „vederii oarbe”, ce apare atunci când cortexul vizual este distrus[6]. Pacienții cu această afecțiune sunt capabili să răspundă la stimuli vizuali, dar, când li se cere să descrie ce „văd”, afirmă că nu au deloc imagini vizuale. În schimb, dacă li se cere, își folosesc instinctele pentru a *ghici* unde se află stimulii vizuali și fac acest lucru cu o precizie remarcabilă. Să luăm cazul pacientului T.N., descris de specialistul în neuroștiințe Lawrence Weiskrantz: deși complet orb – cu alte cuvinte, cu totul lipsit de experiențe vizuale conștiente –, T.N. s-a strecurat cu iscusință printre obstacolele amplasate în calea sa de-a lungul unui coridor. Apoi, când a fost întrebat, a spus că habar n-avea că evita ceva[7]. Această capacitate surprinzătoare este posibilă deoarece calea de la nervul optic la coliculii superiori din trunchiul cerebral a rămas intactă, în pofida absenței cortexului occipital[8].

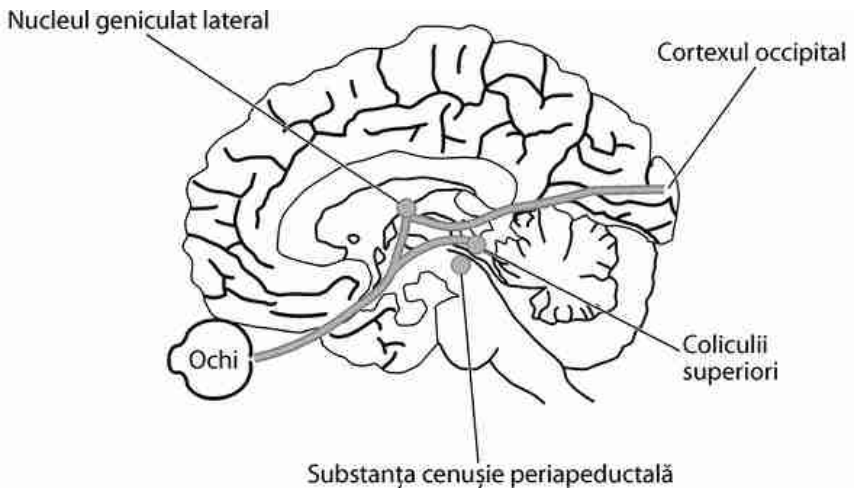


Figura 6. Această figură arată amplasarea coliculilor superiori și a cortexului occipital și conexiunile lor cu ochiul. Conexiuni similare există și pentru celelalte modalități senzoriale. Este

indicată și substanţa cenușie periapeductală (PAG), despre care nu am discutat încă, dar a cărei relevanță va fi evidentă în curând.

Se consideră că existența vederii sugerează că conștientizarea percepției vizuale survine probabil în cortex, și nu în trunchiul cerebral. Se presupune că „nu e nimeni acasă” în trunchiul cerebral: acesta este o mașinărie autonomă ce prelucrează informațiile vizuale în același mod inconștient în care o face o cameră de luat vederi. Acest principiu se aplică și în cazul celorlalte simțuri, fiecare implicând propria zonă de conștientizare în cortex, dar fiecare (cu excepția simțului olfactiv) transmițând și informații coliculilor superiori inconștienți ai trunchiului cerebral.

Fetița înfățișată în figura 5 nu are cortex funcțional. Dacă raționamentul de mai sus este corect, probabil că ea își simte *inconștient* frățiorul așezat în poală, fără generarea unei percepții conștiente a situației. De fapt, ea trebuie să fie incapabilă de orice fel de experiență conștientă. Ea ar fi cam ceea ce filosoful David Chalmers numește „zombi”. Deși din anumite puncte de vedere ea acționează normal, înăuntru e întuneric[*2].

Într-o analiză amplă a tipurilor a ceea ce numim „conștiință”, neurologul Adam Zeman face distincție între două sensuri principale ale termenului: „conștiința ca stare de veghe” și „conștiința ca experiență”[9]. Ulterior, Anton Coenen a elaborat: „Conștiința în primul ei sens (conștiința ca stare de veghe) este, din această perspectivă, o *condiție necesară* pentru conștiința în al doilea sens (conștiința ca experiență sau conștiința fenomenologică)”[10].

Cele două semnificații coincid cu distincția convențională în neurologie dintre „nivelul” cantitativ și „conținutul” calitativ ale conștiinței. Prin urmare, e posibil ca, deși animalele supuse procedurii de decorticare și copiii cu hidroanencefalie sunt în stare de veghe, experiența lor să fie lipsită de conținut. Am văzut mai sus că sunt receptivi și prezintă inițiative comportamentale. Cu toate acestea, putem păstra ipoteza conform căreia cortexul este sediul „conștiinței ca experiență” dacă adăugăm că *a fi* conștient în sensul comportamental de a fi în stare de veghe și a reacționa este semnificativ diferit de *a avea* conștiință în sens fenomenologic – adică a fi subiect al experienței.

În cazul în care noi doi semănăm, în acest punct poate că te simți ușor neliniștit. Fata din figura 5 este conștientă în sensul că este trează și are reacții și inițiază singură acțiuni orientate spre obiective. În același timp, dacă „conștiință” înseamnă să ai experiențe fenomenologice, atunci probabil că ea nu are așa ceva. Ca să folosesc o altă expresie populară în rândul filosofilor, *nu există „ceva cum e” să fii ea*.

Iată ce mi se pare atât de tulburător în privința acestui raționament, chiar dacă sună puțin naiv. Analizându-mi propria situație, a avea reacții și a fi în stare de veghe și a avea o experiență conștientă sunt, mai mult sau mai puțin, același lucru. Din câte știu, eu nu sunt niciodată în stare de veghe și am reacții, dar sunt și inconștient din punct de vedere fenomenologic. Aceste două aspecte merg mână-n mână. De îndată ce mă trezesc, devin conștient de diverse lucruri. De fapt, în opinia mea, mi se pare că conștiința mea interioară îmi *cauzează* reacția la stimuli externi, cel puțin într-o anumită măsură. De regulă, când observ ceva – adică atunci când

conștientizez acest ceva –, am și o reacție din punct de vedere intențional. Presupun că și în cazul tău e la fel.

Acesta este motivul pentru care ne facem o idee cu privire la conștiința pacienților cu antecedente neurologice în funcție de reacțiile lor. Ce altceva mai putem face? În practica clinică, facem distincție între stări precum coma, starea vegetativă și starea de veghe cu reacții complete cu ajutorul scalei de comă Glasgow în 15 puncte. Această scală este alcătuită din teste ale răspunsului ocular al pacienților, ale răspunsurilor lor verbale la întrebări și ale răspunsului lor motor la instrucțiuni și (dacă este necesar) la durere[11]. Dacă pacientul are răspunsuri complete, îl considerăm conștient și îl tratăm ca atare. Nu ne punem problema că ar putea răspunde *ca și cum* ar fi conștient când, de fapt, este zombi. Problema apare în sens invers: neurologii trebuie să verifice dacă există posibilitatea ca pacientul să nu aibă reacţii la nivel exterior, dar să fie conștient la nivel interior – cum se întâmplă, de pildă, în cazuri cu pierderea totală a motricităţii precum pseudocoma.

Ce putem conchide de aici? E de netăgăduit că problema filosofică a altor minți ridică semne de întrebare cu privire la ce putem ști despre stările subiective ale animalelor și ale altor oameni din comportamentul lor, cam tot așa cum minciunile sau prefăcătoriile ne pot induce în eroare cu privire la stările interioare ale unei persoane. Dar problema altor minți *nu* stabilește că experiența interioară și reacţia exterioară sunt de fapt independente una de alta. Întrucât fiecare simte prezența unei legături puternice în cazul său individual, sarcina probei le revine cu siguranță celor care vor să pretindă că comportamentul de răspuns emoțional nu implică experiența fenomenologică. Iată ce prevăd în mod normal regulile științei.

Aceste afirmații sunt valabile mai ales când avem de-a face cu ființe umane despre care, indiferent de ce afecțiuni neurologice ar mai suferi, putem presupune că își generează comportamentul din starea de veghe într-un mod oarecum obișnuit – iar prin asta mă refer la funcționarea normală a regiunilor cerebrale fără leziuni, și nu la minciuni și prefăcătorie și nici la tipul de născociri viclene pe care ne-am putea aștepta să le găsim la niște roboți filosofici imaginari.

Implicațiile pentru etica medicală sunt considerabile. O colegă de-a mea, psihiatră, al cărei copil are hidroanencefalie mi-a mărturisit recent o dilemă teribilă cu care s-a confruntat după ce a devenit mamă. Neurochirurgul care îi trata copilul i-a sugerat că o intervenție pentru închiderea fontanelei se putea face fără anestezie deoarece, neavând cortex, copilul era, evident, incapabil să simtă durere. Colega mea nu mi-a spus ce decizie a luat, iar întrebarea ar fi fost prea delicată. Totuși, putem vedea din asemenea cazuri cum unele considerații teoretice aparent abstracte pot duce foarte ușor la erori medicale îngrozitoare. Prin urmare, vreau să subliniez foarte clar: ca să acceptăm că o persoană ce *pare* să fie conștientă *nu este* de fapt, trebuie să cerem argumente extrem de convingătoare. Simpla invocare a îndoielii filosofice nu este suficientă. Avem nevoie de argumente foarte puternice ca să credem că cele două tipuri de conștiință s-au separat la asemenea persoane, întrucât se pare că nouă nu ni se întâmplă niciodată așa ceva.

În prezent, aceste argumente sunt furnizate de teoria conform căreia conștiința survine exclusiv în cortex. E ceva atât de obișnuit să se accepte această idee în domeniul neuroștiințelor behavioriste, încât nu mi-a trecut prin cap să o

pun la îndoială până când, în baza cercetărilor mele despre vise, Allen Braun mi-a atras atenția asupra rolului misterios jucat de trunchiul cerebral. În virtutea teoriei conform căreia conștiința survine exclusiv în cortex, doctorii hotărăsc dacă își tratează sau nu pacienții ca pe niște ființe capabile de simţire. De exemplu, unii îngrijitori se ocupă de copii cu hidroanencefalie în condiții de neglijare emoțională severă în baza supoziției că cei mici sunt „vegetativi”. Dacă teoria este corectă, atunci nu avem nicio problemă: nu e vorba de fapt de neglijare deoarece copiii sunt asemenea zombilor filosofici. Ei par să aibă sentimente, dar de fapt nu au. Impresia că au este o iluzie produsă de problema altor minți. Lăsându-ne înşelaţi de comportamentul lor exterior, ne imaginăm că au o oarecare natură interioară[12].

Așadar, ce este teoria conform căreia conștiința survine exclusiv în cortex și cât e de convingătoare?

Primul lucru de observat este că ea a început să se dezvolte foarte devreme. Observația curentă conform căreia conștiința noastră constă în principal din imagini perceptive ale unor evenimente ce se desfășoară în jurul nostru sugerează că fluxul conștiinței spre interior trece prin simțuri. Fără îndoială, această opinie de bun-simț s-a perpetuat de când am început să ne gândim la asemenea chestiuni. În secolele al XVII-lea și al XVIII-lea ea a dat naștere filosofiilor „empiriste” ale lui John Locke și David Hume. Ei au formulat teoria că mintea – care la început e o tabula rasa – își capătă toate caracteristicile specifice prin impresiile lăsate de vibrațiile senzoriale. Se presupunea că impresiile se asociază unele cu altele prin îmbinări regulate de diferite tipuri[13] pentru a ne produce

imaginile mnezice ale obiectelor care, la rândul lor, devin pietrele de temelie ale unor idei mai abstracte. Vibrațiile senzoriale ulterioare stimulează aceste imagini asamblate în prim-planul conștiinței, astfel încât ceea ce trăim nu sunt senzații brute, ci *ceea ce am învățat* despre lume.

Modul în care ideile devin conștiente ca răspuns la un stimul extern a fost numit *apercepție* (care înseamnă aproximativ percepţia prezentului prin prisma experiențelor trecute)[14]. S-a spus că procesele cognitive precum imaginile mentale folosite în gândire implică aproximativ același proces în sens invers: activări generate intern ale imaginilor mnezice, rearanjate adecvat (și mai slabe decât cele generate extern).

Deși era speculativă, această idee filosofică despre minte a devenit harta după care s-au ghidat primii neurologi. Când pionierii din secolul al XIX-lea ai neuropsihologiei moderne au căutat să stabilească corelativele neurale ale acestor procese, au observat că organele de simţ erau conectate la cortex și au presupus că „vibrațiile” senzoriale aveau loc în acești nervi de legătură. Se întâmplă că nu au trecut cu vederea faptul că organele de simţ erau conectate și cu nucleii subcorticali. Totuși, au presupus că marele depozit de imagini mnezice care constituie cunoștințele noastre despre lume se află în cortex, deoarece acesta conține incomparabil mai mulți neuroni. Astfel, s-a presupus că „apercepția” și „ideile” aferente care generează activitatea mintală în sine sunt fenomene corticale. Theodor Meynert, marea autoritate a epocii în materie de neuroanatomie, s-a exprimat astfel:

> Principala funcție a organului central este să transmită faptul existenței către un eu care se modelează treptat în

> fluxul creierului [...]. Dacă privim cortexul drept un organ ce funcționează ca un întreg, atunci se poate spune numai că el este util în procesele minții [...]. E imposibil și nenecesar să analizăm cortexul mai mult de-atât[15].

Cortexul este, desigur, impresionant de mare la oameni[16]. Neurologul Alfred Walter Campbell, specialist în neuroanatomie, a rezumat acest curent de gândire dominant în cadrul întâlnirii anuale a Asociației de Medicină și Psihologie din Londra, în 1904:

> Văzut în ansamblu, creierul uman include două tipuri de centri, controlând ceea ce putem numi funcții „evoluționare primare" și, respectiv, „superioare"; primele sunt cele comune la toate animalele și esențiale pentru supraviețuire, adică centrii pentru mișcare și senzații comune și speciale; cele din urmă sunt funcțiile psihice complexe care îl fac pe om să se ridice deasupra tuturor celorlalte ființe[17].

Este important să recunoaștem că, din această perspectivă, mintea constă în întregime din imagini mnezice reflectând experiențele trecute ale lumii exterioare. Senzațiile noi pur și simplu stimulează aceste imagini și asocierile lor în conștiință. Prin urmare, vibrațiile ce intră prin intermediul organelor de simţ sunt evenimente prepsihice – *declanșatori* ai activității mintale –, ele nefiind evenimente psihice. Aceeași afirmație este valabilă şi în cazul funcției nervilor ce conectează cortexul cu restul corpului: căile lor nu au nimic „psihic", ele pur și simplu eliberează *outputurile* activității psihice. Activitatea mintală propriu-zisă nu poate avea loc decât în cortex, unde se află imaginile mnezice.

Meynert a descris relația dintre cortex și lumea exterioară după cum urmează:

> Efectele motoare ale conștiinței noastre ce răspunde la lumea exterioară nu sunt rezultatul unor forțe inerente creierului. Creierul, ca o stea fixă, nu radiază căldură: își obține energia susținând toate fenomenele cerebrale din lumea de dincolo de el[18].

Probabil merită să evidențiem și că, întrucât Freud încă nu apăruse, se presupunea că toată activitatea psihică este conștientă. Se considera că „psihic” și „conștient” înseamnă același lucru.

Activitățile experimentale din secolul al XIX-lea care au condus la reprezentarea modernă a conștiinței s-au desfășurat exclusiv în acest cadru filosofic. La sfârșitul anilor 1800, fiziologul Hermann Munk a identificat cortexul occipital drept sediul aspectului mental al vederii (vezi figura 6). A devenit interesat de comportamentul câinilor cu leziuni experimentale la nivelul cortexului occipital. Aceste nefericite animale puteau să vadă, dar se pare că le lipsea „înțelegerea” normală a ceea ce vedeau: de exemplu, nu mai puteau să își recunoască vizual stăpânii sau să-și identifice castronul cu mâncare, deși se uitau la ele și le dădeau ocol, dar le puteau recunoaște cu ajutorul celorlalte simțuri[19]. Munk a numit această afecţiune „cecitate psihică”, pentru a o deosebi de forma comună de orbire cauzată de leziuni la nivelul căilor senzoriale (subcorticale) ducând de la ochi la cortex[20]. În spiritul filosofiei empiriste, a pus semnul egalității între ceea ce numea vedere „psihică” și capacitatea de a activa *imaginile mnezice* vizuale prin asocieri, în opoziţie cu

procesul mecanic al receptării senzațiilor vizuale brute și al declanşării unor reflexe motoare. În consecință, afecţiunea descrisă de Munk este numită acum „agnozie” vizuală – adică lipsa *cunoașterii* vizuale.

Fenomene clinice de exact același tip au fost curând raportate şi în cazul unor oameni – de exemplu, în 1887, de oftalmologul Hermann Wilbrand, care a descris cazul lui Fräulein G., o femeie de 63 de ani care a suferit un accident vascular occipital bilateral[21]:

> Toți din jur o considerau oarbă. Totuși, ea era conștientă că nu era complet oarbă, „pentru că, atunci când oamenii stăteau la căpătâiul meu și vorbeau cu milă de orbirea mea, îmi spuneam: nu pot fi cu totul oarbă pentru că pot să văd fața de masă de acolo, cu margine albastră, întinsă pe masa din infirmerie” [...]. Când această doamnă – altfel foarte inteligentă – s-a ridicat [după ce inițial și-a pierdut cunoștința, din cauza accidentului vascular], și-a dat seama că era într-o stare ciudată în care nu vedea, fiind totuși capabilă să vadă [...]. Și astăzi este tulburată când își amintește prima ieșire din casă după accidentul vascular; cât de diferit și de ciudat i se părea orașul și cât de tulburată și de angoasată s-a simțit când a fost condusă de însoțitoarea ei pentru prima dată pe Jungfernstieg și de la Neuer Wall la Stadthaus și cum, după indicațiile însoțitoarei, clădirile și străzile care îi erau de obicei atât de familiare parcă erau ceva nou. Iată cum își relatează Fräulein G. reacția la afirmațiile femeii care o însoțea la vremea respectivă: „Dacă spui că aici este Jungfernstieg, acolo Neuer Wall, iar acolo Stadthaus, presupun că așa trebuie să fie, deși eu nu le

recunosc [...]. I-am spus medicului meu: «Din afecțiunea mea se poate trage concluzia că vedem mai mult cu creierul decât cu ochii; ochiul nu este decât vehiculul vederii, deoarece eu văd orice lucru în mod absolut clar și lucid, dar nu recunosc lucrul și adesea nu știu ce ar putea fi lucrul pe care îl văd»".

Wilbrand a concluzionat că această pacientă suferise o pierdere a memoriei vizuale (adică o tulburare a recunoașterii, înțelegerii sau cunoașterii vizuale), și nu o simplă orbire. Și a observat că, deși pacienții orbi încă pot genera imagini vizuale în visele lor, întrucât capacitatea lor de a produse imagini mentale este intactă, persoanele oarbe *psihic* precum Fräulein G. nu pot; ele își pierd capacitatea de a visa. Și s-a întrebat cum poți genera halucinații vizuale fără imagini mnezice vizuale[22].

Asemenea observații despre vedere erau generalizate pe atunci şi la alte modalități de percepție. Astfel, ablația cortexului auditiv (din nou, la câini) producea „surzenie mintală", numită astăzi agnozie auditivă, animalele afectate nemaiputând răspunde la sunete cu semnificație dobândită, chiar dacă în mod clar nu erau surde: răspundeau la zgomote brute, dar nu își mai recunoșteau numele când erau chemate. În 1874 neurologul Carl Wernicke a observat ceva similar la pacienții umani, iar astfel a formulat ideea de tulburări ale limbajului dobândit numite „afazie"[23].

La nivel motor, Hugo Liepmann a afirmat că și leziunile corticale produc deficiențe ale aspectului psihic al mișcării, numite „paralizie psihică" (sau „apraxie"). O leziune la nivelul căilor motoare de ieșire a cauzat paralizie fizică, dar o leziune la nivelul centrului cortical pentru imaginile mnezice motoare a cauzat uitarea abilităților de mișcare dobândite (numită

apraxie „chinestezică a membrelor”), iar o leziune la nivelul căilor transcorticale pentru asocierile motoare a cauzat o întrerupere între mișcarea deprinsă și ideile abstracte, cum ar fi semnificația simbolică a gesturilor mâinii (apraxie „ideomotoare”)[24].

Orbirea, surditatea și paralizia subcorticale au fost astfel conceptualizate ca afecțiuni *fizice*, iar efectele vizuale, auditive și motoare ale leziunilor corticale erau *mintale*: formele aperceptive și asociative de agnozie, afazie și apraxie. Aceste din urmă afecțiuni sunt numite astăzi tulburări „neurocognitive”. Astfel, distingerea între aceste două clase de tulburări (subcorticale versus corticale) a coincis cu trasarea graniței disciplinare dintre neurologie și neuropsihologie și rămâne valabilă.

Când contemporanii lui Munk au extirpat complet cortexul la câini – nu doar părțile specializate responsabile cu anumite tipuri senzorimotoare –, animalele nu au intrat în comă sau într-o stare vegetativă, ci s-au comportat ca și cum ar fi fost *fără minte* (în accepțiunea empiristă a termenului). Animalele au devenit amnestice, adică și-au pierdut toate imaginile mnezice și, prin urmare, „înțelegerea”. În consecință, Friedrich Goltz le-a descris drept „nătânge”. Faptul că nu au intrat în comă nu i-a surprins pe primii cercetători deoarece, având în vedere ipotezele teoretice ale vremii, ei se așteptau doar ca acești câini supuși procedurii de decorticare cerebrală să-și piardă toate *cunoștințele* dobândite. Nu era neapărat ca această pierdere să aibă efect şi asupra senzațiilor și reflexelor corporale, despre care se presupunea că sunt subcorticale. Ideea că viața psihică implică doar imagini mnezice nu era controversată. La vremea respectivă, controversele vizau alte chestiuni, cum ar fi cât de

precis putea fi localizată fiecare funcție psihică în regiuni specifice ale cortexului. Nimeni nu se îndoia că toate funcțiile de acest fel sunt corticale.

Așa că, atunci când Meynert a rezumat tot tabloul ce se contura în celebra lui carte din 1884 *Psychiatry*[25], a identificat mintea cu *totalitatea imaginilor mnezice ale obiectelor* produse prin proiecția periferiei senzorimotoare în cortex plus asocierile transcorticale dintre ele și imaginile mnezice care constituiau ideile abstracte. În mod firesc, a localizat toate aceste imagini, asocieri și idei în cortex – conform citatului de mai sus. L-a numit partea „voluntară” a creierului și a pretins că are legături directe cu periferia trupului prin nervii senzoriali și motori. Adică, în opinia sa, cortexul era conectat cu corpul *independent* de substanța cenușie subcorticală. A recunoscut că partea subcorticală a creierului era legată de asemenea de cortex și de periferia corpului, prin propriile căi distincte, dar le-a descris drept „reflexe”. Aici avem cei „doi creieri” pe care i-am menționat mai sus, care explică comportamentul copiilor cu hidroanencefalie.

În acest fel, cortexul a devenit organul minții – mintea concepută drept *conștiință a imaginilor mnezice* –, iar creierul subcortical a devenit fără minte. Totul se reducea la ideea că „mintea” este constituită în întregime din experiențe trecute, care lasă urme asociate între ele pentru a forma imagini concrete și idei abstracte. Restul, părțile periferice senzorimotoare și părțile înnăscute și subcorticale, inclusiv părțile care transmit impresii din *interiorul* corpului, erau considerate exclusiv reflexe. Și așa, deși pare ciudat, distincția filosofică dintre minte și corp a ajuns să coincidă cu distincția anatomică dintre cortex și subcortex.

În anii 1960, Norman Geschwind, marele pionier al neurologiei behavioriste din America, a reluat cu entuziasm aceste idei clasice[26]. Renașterea asociaționismului empirist în neurologie a coincis cu „revoluția cognitivă” din psihologie. Schemele cu informații ale specialiștilor în științe cognitive semănau foarte mult cu diagramele folosite de specialiștii germani clasici în neurologie pentru a ilustra relațiile funcționale dintre centrii corticali ce conțin imagini mnezice de diferite tipuri. Dar – după cum vom vedea – cognitivismul modern a condus și la conștientizarea (consacrată până în anii 1990) a faptului că multe funcții psihice (inclusiv percepția și memoria) *nu sunt totuşi conștiente.*

Cam în perioada în care Geschwind reintegra în neuropsihologie teoria clasică potrivit căreia conștiința apare exclusiv în cortex, pionierii (precum Paul MacLean și James Olds) a ceea ce avea să devină neuroștiința afectivă acumulau observații care au dezvăluit că funcții psihice îndeplineau și numeroși nuclei subcorticali ce conectau creierul cu interiorul corpului, funcții precum *imboldul* motivațional și *trăirea* emoțională.

În pofida acestor evoluții paralele în neuroștiințele cognitive și afective, pe care le voi elabora mai jos, până la urmă această moștenire intelectuală – încercarea neurologilor din secolul al XIX-lea de a confirma teoriile filosofilor din secolul al XVIII-lea – e cea pe care se clădește prezumția neurologilor de astăzi conform căreia copiii cu hidroanencefalie sunt fără minte. Ei fac această prezumţie, deși recunosc că echivalarea de altădată a „psihicului” cu conștiința și cu cortexul nu mai este valabilă: că nu întreaga activitate psihică este conștientă și nu toată conștiința este corticală. Cu alte cuvinte, insistența asupra ideii

că această echivalare este încă valabilă, că „conștiința ca experiență” e cu necesitate corticală are la bază o inerție a teoriei, și nu nişte dovezi științifice.

După cum poate ţi-ai dat seama, eu nu consider că teoria conform căreia conștiința survine exclusiv în cortex este valabilă. De fapt, aș spune chiar că animalele și oamenii pot fi conștienți chiar dacă nu au deloc cortex. Eu cred că *există* „ceva cum e” să fii un copil cu hidroanencefalie.

Desigur, unul dintre motivele pentru care e atât de dificil să știm ce trăiri au pacienții cu hidroanencefalie – dacă au vreo experiență interioară – este că ei nu pot să vorbească. Limbajul chiar este o funcție corticală, așa că nu ne putem aștepta ca oamenii care nu o au să ne furnizeze rapoarte verbale introspective. Ei nu ne pot furniza tipul de dovadă subiectivă care de obicei ne face să credem că alți oameni sunt conștienți, în pofida problemei altor minți[27]. Afirmația e valabilă și în cazul animalelor. Totuși, există şi oameni care şi-au pierdut o porţiune substanțială din cortex, dar nu și capacitatea de a vorbi. În aceste cazuri, pur și simplu îi putem *întreba* cum stă treaba.

Neurochirurgii Wilder Penfield și Herbert Jasper au extirpat porţiuni mari din cortex (chiar emisfere întregi) la 750 de pacienți umani sub anestezie locală, în principal pentru tratarea epilepsiei. Au observat că asemenea intervenții au efecte limitate asupra conștiinței autoraportate, chiar în momentul în care cortexul este îndepărtat. (Intervențiile chirurgicale pe creier se fac adesea sub anestezie locală, astfel încât pacientul să poată raporta efectele a ceea ce face chirurgul.) Penfield și Jasper au concluzionat că rezecțiile

corticale nu întrerup existența conştientă, ci doar îi privează pe pacienți de „anumite forme de informare”[28]. Eu am asistat la multe asemenea intervenții, în timpul cărora de obicei rolul meu e să evaluez efectele stimulării electrice a ariilor din cortex responsabile cu memoria și limbajul înainte de intervenția chirurgului. Am observat exact ceea ce au raportat Penfield și Jasper.

Dar nu tot cortexul este la fel. Regiunile cortexului care primesc inputuri de la fiecare dintre simțurile noastre speciale generează „qualia” fenomenale asociate cu simțurile respective (culoare pentru vedere, ton pentru auz și așa mai departe). Totuși, conform opiniei standard, senzațiile brute pot fi declarate numai după ce au fost accesate de un tip general de conștiință care constituie „eul” conştient[29].

Sunt utilizate diferite terminologii pentru a descrie relația dintre conștiința „fenomenală” (de exemplu, simpla vedere și auzire) și conștiința „reflexivă” (*știi* că vezi și auzi). Dar toate transmit aceeași idee de bază: ești mai mult decât diversele forme de informații senzoriale pe care le prelucrezi. Așa se explică faptul că, deși unii pacienți cu leziuni la nivelul cortexului vizual primar sunt orbi, iar unii cu leziuni la nivelul cortexului auditiv sunt surzi, ei nu și-au pierdut simţul *sinelui*. Acest sine conștient ghicește ce vezi când nu ai qualia vizuale în cazul vederii oarbe, de exemplu. Asemenea pacienți pierd doar „anumite forme de informare”. Așadar, întrebarea arzătoare este: ce învățăm de la pacienții care și-au pierdut porțiunile de cortex despre care se spune că sunt responsabile de ipseitate?

Există trei locuri ce candidează pentru această funcție. Primul este *cortexul insular*, care este specializat în conștientizarea interoceptivă și despre care se consideră în

general că generează sentimentele ce constituie „sinele” conştient[30]. A doua este *cortexul prefrontal dorsolateral*, care formează o suprastructură peste toate celelalte părți ale creierului și despre care se consideră în general că permite „gânduri de ordin superior”, inclusiv conștientizarea sentimentelor[31]. A treia este *cortexul cingular anterior*, care se activează în mai toate experimentele de imagistică cognitivă a creierului și despre care se spune că mediază aspecte precum „prelucrarea informațiilor despre sine” și „voința”[32].

Să luăm aceste trei regiuni pe rând.

Iată un fragment dintr-un interviu luat de Antonio Damasio unui pacient al cărui cortex insular a fost distrus complet de o infecție virală numită encefalită cu *herpes simplex*[33]:

Întrebare: Ai un simţ al sinelui?
Răspuns: Da, am.
Î: Dar dacă ți-aș spune că nu ești aici acum?
R: Aș spune că ești orb sau surd.
Î: Crezi că alți oameni îți pot controla gândurile?
R: Nu.
Î: Și de ce crezi că nu e posibil?
R: Îți controlezi singur mintea, să sperăm.
Î: Dacă ți-aș spune că mintea ta este mintea altuia?
R: Când a avut loc transplantul, adică transplantul de creier?
Î: Dar dacă ți-aș spune că te cunosc mai bine decât te cunoști tu pe tine?
R: Aș crede că te înșeli.
Î: Dar dacă ți-aș spune că ești conștient că eu sunt conștient?
R: Aș spune că ai dreptate.
Î: Ești conștient că eu sunt conștient?

R: Sunt conștient că ești conștient că sunt conștient.

Acest pacient, numit „B.”, a fost studiat de echipa lui Damasio o perioadă de 27 de ani (s-a îmbolnăvit la 48 de ani și a murit la 73 de ani). În studiul de caz publicat, echipa a raportat date neuropsihologice detaliate referitoare la capacitatea lui de a simți și de a răspunde emoțional. Iată concluziile principale:

> Pacientul B., al cărui cortex insular a fost distrus în întregime, avea senzații corporale și trăiri emoționale. Raporta că simțea durere, plăcere, mâncărime, gâdilituri, fericire, tristețe, frică, iritare, afecțiune și compasiune și se comporta în moduri în acord cu ceea ce simţea în momentele respective. Raporta și senzații de foame, de sete și dorința de a excreta și avea comportamentele aferente. Tânjea să se joace – de exemplu, să joace dame –, să-i viziteze pe alții, să se plimbe și manifesta o plăcere evidentă când se implica în asemenea activități, precum și dezamăgire și chiar iritare când nu mai avea acces la astfel de oportunități. […] Dată fiind sărăcirea imaginației sale, existența pacientului B. era o reacție „afectivă” practic continuă la propriile stări fizice și la cerințele modeste ale lumii din jurul său, care nu erau moderate de mecanisme de control cognitiv de ordin superior […]. Conform lui Craig (2011), semnul ce confirmă conștiinţa de sine este capacitatea de a te recunoaște când te vezi într-o oglindă, capacitate care, în cuvintele sale, „nu poate fi furnizată decât de o reprezentare neuronală a sinelui funcțională, validă din punct de vedere emoțional”. Pacientul B. a trecut acest test în mod constant și repetat. Pe scurt, aceste concluzii

> contrazic ipoteza conform căreia conștiința de sine la oameni, alături de capacitatea de a simți, ar depinde în întregime de cortexul insular și, în mod specific, de treimea sa anterioară (Craig, 2009, 2011). În lipsa cortexului insular, trebuie să luăm în considerare alternative neuroanatomice ca să explicăm fundamentul capacităților senzoriale și al facultății lui B. de a simți[34].

Bud Craig și alții afirmă că trăirile afective ce constituie sinele sunt generate în cortexul insular. Și totuși, Damasio a descoperit că, de fapt, pacientul B. era, ca și șobolanii supuși procedurii de decorticare, *mai* emoțional după ce a suferit leziunile la nivelul cortexului decât înainte. Predicția că asemenea pacienți ar trebui să-și piardă „prezența” conștientă este în mod clar invalidată.

Același lucru se întâmplă și la pacienții umani cu leziuni la nivelul celei de-a doua regiuni analizate aici: cortexul *prefrontal*, care constituie o suprastructură deasupra tuturor celorlalte părți ale creierului. Impresionabilitatea pacienților cu leziuni la nivelul cortexului prefrontal este larg recunoscută, constituind o trăsătură esențială în „sindromul de lob frontal”. Acest lucru fusese deja observat în 1868, în celebrul caz a lui Phineas Gage, după ce o bară de fier îi străpunsese craniul într-un accident de muncă:

> Echilibrul, ca să spunem așa, dintre facultățile sale intelectuale și înclinațiile lui animalice pare să fi fost distrus. Este schimbător, nepoliticos, dedându-se uneori la cele mai ordinare obscenități (ceea ce nu îi stătea în obicei înainte), manifestând prea puțin respect față de semenii lui,

respingând cumpătarea sau sfaturile când intră în conflict cu dorințele sale[35].

Unii dintre principalii specialiști în neuroștiințe care studiază emoțiile, precum Joseph LeDoux, consideră că sentimentele *se ivesc* în sensul propriu în cortexul prefrontal dorsolateral[36]. Conform acestei opinii, precursorii subcorticali ai sentimentelor sunt în întregime inconștienți până când sunt „etichetați” în memoria funcțională conștientă[37]. Pentru acești teoreticieni, emoția este doar o altă formă de cogniție – de fapt, o formă de cogniție reflexivă, mai degrabă abstractă. Dar, dacă au dreptate, atunci de ce pacienții a căror memorie funcțională este în mare parte distrusă manifestă atât de multă emoție? Phineas Gage este doar cel mai celebru caz dintr-o literatură consistentă. Ce ar putea să adauge la toată această imagine conștiința reflexivă, dacă forma prereflexivă a simțirii este atât de viu exprimată în comportamentul acestor pacienți?

Trebuie să recunoaștem că majoritatea susținătorilor teoriei conform căreia conștiința survine exclusiv în cortex nu sunt interesați în mod special de sentimente. Ei se axează în schimb pe contribuția lobilor prefrontali la *gândirea* de ordin superior. De asemenea, ipotezele ce rezultă din aceste teorii mai generale pot fi testate imediat în cazul unor oameni cu leziuni la nivelul lobilor prefrontali. Leziunile prefrontale totale sunt foarte rare – dar nu inexistente.

Pacientul meu „W.” avea 48 de ani când l-am examinat prima dată. A obținut scorul Glasgow maxim, ceea ce înseamnă că era complet conștient și avea răspunsuri complete. La vârsta de 13 ani suferise un anevrism cerebral rupt, ce a necesitat o intervenție în care lobii frontali au fost retractați pentru a

acoperi vasul de sânge vulnerabil și, astfel, a preveni altă hemoragie. Din păcate, această intervenție a cauzat o infecție cerebrală cronică, presupunând multiple intervenții suplimentare ce au dus în cele din urmă la distrugerea completă a lobilor prefrontali de ambele părți (vezi figura 7).

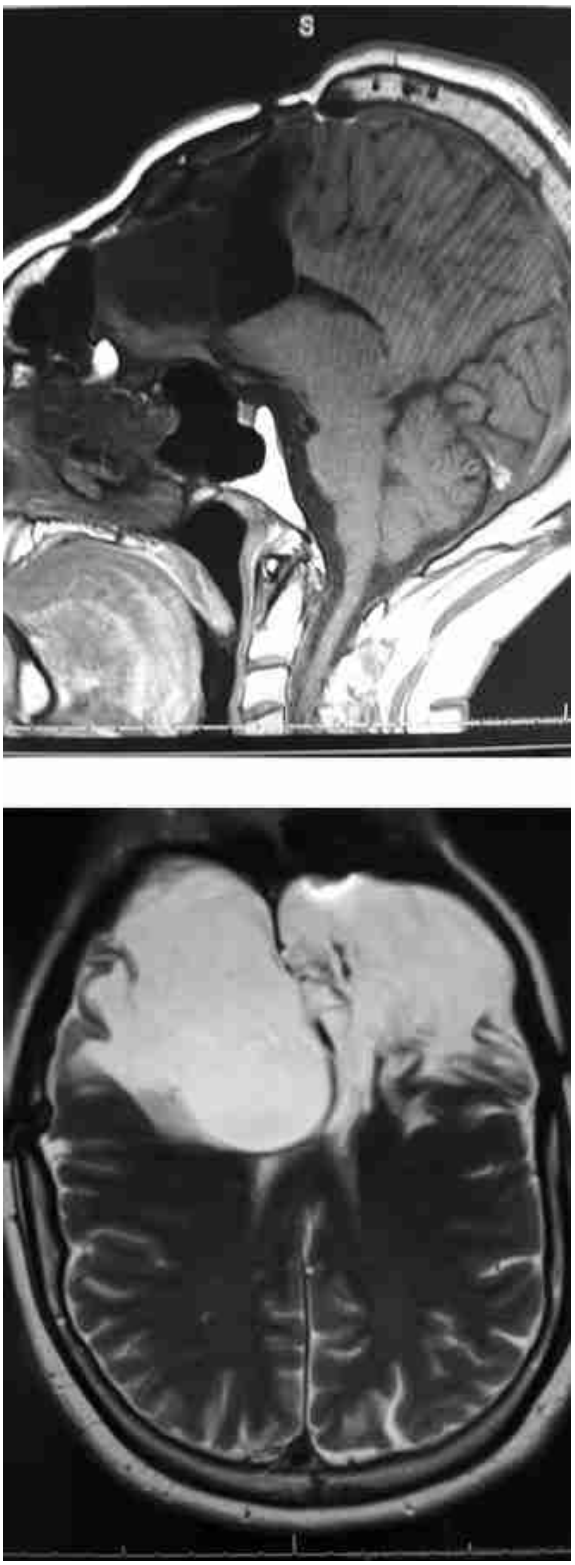


Figura 7. RMN cerebral al pacientului W., ce arată distrugerea completă a lobilor prefrontali de ambele părți

Din fericire, regiunea cerebrală responsabilă cu limbajul nu a fost afectată. Iată o parte din conversația noastră:

Întrebare: Ești conștient de gândurile tale?
Răspuns: Da, sigur că da.
Î: Pentru a confirma acest lucru, te rog să rezolvi o problemă în care ți se cere să-ți imaginezi în mod conștient o situație în minte.
R: OK.
Î: Imaginează-ți că ai doi câini și un pui de găină.
R: OK.
Î: Îi vezi cu ochii minții?
R: Da.
Î: Acum spune-mi câte picioare vezi în total?
R: Opt.
Î: Opt?
R: Da, câinii au mâncat puiul.

Poanta a fost spusă cu un rânjet răutăcios. Poate că nu e cea mai bună glumă din lume, dar constituie o dovadă convingătoare că, folosind o expresie colocvială, era cineva acasă. Pacientul W. a afirmat că este conștient, cum a afirmat și pacientul B. al lui Damasio, iar în lipsa unui argument mai convingător decât scepticismul filosofic radical, eu înclin să-i cred pe cuvânt.

A treia și ultima regiune corticală despre care se presupune că are o relație specială cu „eul” conştient este *girusul cingular anterior*. Pacienții cu leziuni bilaterale totale în această regiune sunt relativ des întâlniţi, parțial din cauză că această regiune a cortexului este o țintă obișnuită a intervențiilor chirurgicale în

afecțiunile psihiatrice, precum tulburarea obsesiv-compulsivă. (Regiunea cingulară anterioară și-a atras toată această atenție pe cap, cum s-ar spune, tocmai din cauza asocierii sale cu „prelucrarea informațiilor legate de sine”).

În perioada postoperatorie acută, unii dintre acești pacienți nu mai pot face diferența între fantezie și realitate. Charles Whitty și Walpole Lewin au descris un exemplu frapant (Cazul 1, în lucrarea lor):

> Ca răspuns la întrebarea referitoare la ce a făcut în ziua respectivă, pacientul a spus: „Eu am stat la ceai cu soția mea”. Fără alte comentarii din partea intervievatorului, a continuat: „A, de fapt eu n-am stat la ceai. Ea n-a fost azi pe-aici. Dar, imediat ce eu închid ochii, scena asta îmi apare într-un mod atât de viu. Eu văd ceștile și farfurioarele și o aud cum toarnă ceaiul. Apoi, cum eu ridic ceașca să iau o gură, eu mă trezesc și nu mai e nimeni”.
> Întrebare: Deci dormi mult?
> Răspuns: Nu, eu nu dorm – e un fel de vis în stare de veghe [...], chiar cu ochii deschiși uneori [...]. Parcă nu e niciun control asupra gândurilor mele, se iscă singure – pare atât de viu. Jumătate din timp eu nu sunt sigur dacă eu doar m-am gândit sau dacă chiar s-a întâmplat[38].

Observați folosirea frecventă a lui „eu”. Whitty și Lewin bănuiau că era posibil ca pacientul lor să sufere de crize complexe, dar în multe alte cazuri cu leziuni bilaterale ale girusului cingular anterior se raportează experiențe similare. Dau mai jos exemplul unei paciente de-ale mele[39]. E vorba de o femeie de 44 de ani care a suferit o hemoragie subarahnoidă. A

raportat următoarele:

> E ca și cum gândirea mea devine reală – când mă gândesc la ceva, imediat văd lucrul acela întâmplându-se aievea în fața ochilor, apoi devin și foarte confuză și uneori nu știu ce s-a întâmplat cu adevărat și ce a fost doar un gând.
>
> Ea a dat un exemplu:
>
> *Pacienta*: Stăteam în pat și mă gândeam, apoi s-a întâmplat cumva pur și simplu că soțul meu [decedat] era acolo și-mi vorbea. Apoi m-am dus să le fac baie copiilor și apoi, brusc, am deschis ochii și „Unde mă aflu?" – și sunt singură!
> *Eu*: Adormiseși?
> *Pacienta*: Nu cred. E ca și cum gândurile mele numai ce s-au transformat în realitate.

Conștiința este evident anormală în aceste cazuri, dar nu asta vreau să arăt. Nu contestă nimeni implicarea cortexului în procesarea conștientă. Ceea ce nu se poate demonstra e opinia că sinele conștient este *generat* acolo.

Observăm comportamente energice, intenționate, de răspuns și emoționale la oameni și animale care nu mai au deloc cortex. Introspecția sugerează că conștiința este intim implicată în aceste comportamente în cazul meu și al tău (prezumtiv normale). Așadar, în pofida problemei altor minți, am avea nevoie de un motiv temeinic să credem că ea nu e implicată și în aceste cazuri. Dar, când căutăm asemenea motive temeinice, găsim doar povara istoriei mediului academic, un model preneurologic al minții care s-a întâmplat

să fie folosit ca șablon de primii exploratori ai creierului. Astăzi, foarte puțin oameni de știință ar susține versiunea empirismului lui Hume sau ideile lui Meynert din volumul *Psychiatry*. Ba chiar opiniile lui Meynert sunt în mare parte luate în râs şi considerate o „mitologie a creierului”[40]. Și totuși, dogma conform căreia cortexul este organul minții a devenit ipoteza fondatoare a unui întreg domeniu al medicinei.

Bineînțeles, cortexul este implicat în multe funcții cognitive, inclusiv limbajul. Prin urmare, nu putem spera să primim rapoarte verbale introspective de la oameni la care este absent în totalitate. Dar avem mărturii nemijlocite de la pacienți ale căror afecțiuni le permit să facă relatări verbale reflexive, chiar dacă le lipsesc părțile specifice ale cortexului despre care se presupune că dau naștere conștiinței de sine generale. De nenumărate ori, acești pacienți afirmă că sunt conștienți și își declară „existența” introspectivă. Oare mint? Dacă da, de ce? Și oare ce-o fi însemnând afirmaţia că o persoană fără un sine minte cu privire la propria ipseitate? Neuroștiința cognitivă e aproape incoerentă în acest punct – semn clar că a luat-o într-o direcție greșită.

Eu consider că, în baza dovezilor existente, teoria conform căreia conștiința survine exclusiv în cortex nu se confirmă. Nu există motive temeinice să credem că cortexul dă naștere existenței conştiente, în sensul în care o trăim și eu, și tu în mod obișnuit, dar sunt multe motive temeinice să concluzionăm că această afirmație nu e valabilă. Trebuie să căutăm în altă parte sursa ființei noastre.

*1 E o ironie că rezultatul unor astfel de experimente a contribuit în cele din urmă la schimbarea atitudinilor

noastre față de etica unor asemenea studii.

*2 Noțiunea filosofică de zombi e diferită de cea hollywoodiană: aici e vorba de o creatură umanoidă imaginară ce acționează din toate punctele de vedere *ca și cum* ar fi conștientă și, din acest motiv, este nediferențiabilă de o persoană normală, dar de fapt îi lipsește dimensiunea interioară a experienței subiective.

4

Care este obiectul experienței?

Trebuie să fii conștient de ceea ce percepi și înveți ca să percepi și să înveți? Poate că bunul-simț spune da. Filosofii empiriști au spus da. Dar răspunsul este nu.

Ideile empiriștilor au condus la concepţia clasică din neuroanatomie conform căreia percepția (și engramele derivate din ea) este ingredientul de bază al conștiinței. Dar dovezile științifice care arată că *nu suntem conștienți de majoritatea lucrurilor pe care le percepem și le învățăm* sunt acum copleșitoare. Percepția și memoria nu sunt funcții cerebrale inerent conștiente. În acest sens, bunul-simț s-a înşelat. Se dovedește că tot ce face mintea (cu excepția unui singur lucru, după cum vom vedea) poate fi făcut destul de bine şi în mod inconștient.

E ciudat că această idee îi poate fi atribuită lui Sigmund Freud.

Deși toată viața a susținut teoria conform căreia conștiința survine exclusiv în cortex, Freud a fost unul dintre primii specialiști în neuroștiințe care au pus sub semnul întrebării distincția făcută de Meynert, profesorul lui, între părțile „psihice” și cele „nonpsihice” ale creierului[1]. După ce a analizat

datele disponibile, a concluzionat în 1891 că opiniile lui „Munk și ale altor cercetători care pleacă de la ideile lui Meynert [...] nu mai pot fi susținute”[2]. Mai exact, nu a găsit nicio deosebire anatomică clară între căile „voluntare” și cele „reflexe” postulate de Meynert. De asemenea, a demonstrat (numărând fibrele nervoase implicate) că imaginile mnezice corticale erau generate numai după o serie de legături intermediare dintre ele și periferia senzorială. Aceste legături se aflau în regiunile presupuse a fi *nonpsihice* ale creierului și reduceau volumul de informații transmise în fiecare etapă. Freud a dedus că aceste legături subcorticale trebuie să aibă vreun efect asupra informațiilor senzoriale pe care le prelucrează. De observat metafora lui uimitoare, concepută într-o eră predigitală:

> Știm doar că fibrele care ajung la cortexul cerebral după trecerea prin țesuturile cenușii [cele subcorticale] încă păstrează o anumită relație cu periferia corpului, dar nu mai pot transmite o imagine care să-i semene din punct de vedere topologic. Ele conțin periferia corpului în același fel în care [...] un poem conține alfabetul, într-o rearanjare totală, servind unor scopuri diferite, cu numeroase legături între elementele topologice individuale, prin care unele dintre ele pot fi redate de mai multe ori, iar altele deloc[3].

În cazul în care cortexul nu este conectat direct la periferia corpului, ci, mai degrabă, prin legături subcorticale intermediare, imaginile mnezice depozitate în cortex nu pot fi proiecții propriu-zise ale lumii exterioare. Ele sunt probabil produsele finale ale unei prelucrări a informațiilor în mai multe etape. Întrucât această prelucrare culminează cu

imaginile mnezice corticale, partea subcorticală a prelucrării trebuie să genereze, în vreun sens, *versiuni preliminare* ale imaginilor mnezice. Prin urmare, legăturile subcorticale constituie probabil o parte din procesul mintal pe care îl numim „apercepție”. Freud afirma că nu are sens să facem o distincţie artificială între părțile subcorticale și cele corticale ale prelucrării informațiilor și să pretindem că numai produsul final este „psihic”:

> Se justifică să luăm o fibră nervoasă, care pe toată lungimea traseului ei este doar o structură fiziologică supusă modificărilor fiziologice, să-i cufundăm capătul în psihic și să furnizăm acestui capăt o idee sau o imagine mnezică?[4]

Întrucât Freud a presupus, ca toată lumea din perioada aceea, că numai procesele corticale sunt *conștiente*, acest lucru l-a determinat să susțină ideea că percepția și învățarea trebuie să includă etape preliminare *inconștiente* care sunt la fel de „psihice” ca şi cele conștiente. Adică engramele inconștiente (subcorticale) trebuie să fie la fel de psihice ca şi cele corticale: și ele fac parte din funcția pe care o numim „memorie”, deși le lipsește conștiința. Freud a concluzionat într-o scrisoare către Wilhelm Fliess: „În esență nouă în teoria mea este teza că memoria e prezentă nu o dată, ci de mai multe ori, că e afișată în diferite tipuri de indicii”[5].

A identificat cinci etape succesive în prelucrarea informațiilor: „percepția”, „urma perceptivă”, „etapa inconștientă”, „etapa preconștientă” și „etapa conștientă” (vezi figura 8). Diferența crucială dintre etape consta nu în faptul că etapa inconștientă era corporală, iar cea preconștientă era

psihică, ci că tipul preconștient de procesare a memoriei putea fi refăcut în conștiință, dar nu și tipul inconștient[6]. Cu alte cuvinte, *doar o parte a psihicului este conștientă.*

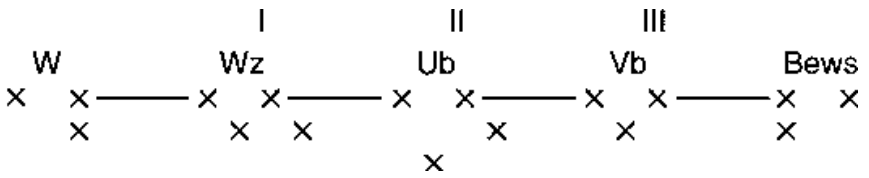


Figura 8. Prima diagramă a sistemelor de memorie realizată de Freud, ataşată scrisorii sale către Wilhelm Fliess (*W* = percepție, *Wz* = urma perceptivă, *Ub* = etapa inconștientă, *Vb* = etapa preconștientă, *Bews* = etapa conștientă)

În 1895 Freud a făcut o observație similară despre căile nervoase care pleacă din *interiorul* corpului. Nici în acest caz, a afirmat el, nu are sens să spunem că informațiile corporale viscerale devin „psihice” numai când ajung la cortex. Întrucât aceste informații (care transmit senzații precum foamea sau setea) solicită psihicului să lucreze ca urmare a conexiunii cu corpul, trebuie să înglobăm și aceste solicitări în imaginea noastră despre psihic. Chiar dacă nu îți conștientizezi „pulsiunile” biologice, cum le numea Freud, cu siguranță ele fac parte din psihicul tău. Amintește-ți din capitolul 2 că Freud a sugerat chiar că solicitările corpului furnizează „motorul mecanismului psihic” și că prozencefalul (al cărui cortex reprezintă lumea exterioară) este doar un „ganglion simpatic”. Cu alte cuvinte, el a concluzionat că atât imaginile mnezice conștiente, cât și cele inconștiente sunt formate „în simpatie” cu solicitările din partea corpului – că ne reprezentăm și învățăm

despre lumea exterioară numai prin intermediul percepției, deoarece acolo trebuie să ne satisfacem nevoile biologice.

La această perspectivă asupra creierului știința modernă adaugă multe detalii despre care Freud nu avea nicio idee și îl corectăm în unele puncte importante. Chiar și așa, concluzia lui fundamentală este în general acceptată astăzi în neuroștiințe: creierul îndeplinește o mare diversitate de funcții *psihice* care nu intră în conștiință. Titlul unei celebre treceri în revistă a literaturii relevante semnate de John Kihlstrom, un specialist modern în științele cognitive, e cât se poate de sugestiv: există într-adevăr „percepție fără conștientizarea a ceea ce este perceput, învățare fără conștientizarea a ceea ce este învățat”[7]. O altă analiză bine-cunoscută, scrisă de John Bargh și Tanya Chartrand, rezumă perspectiva pe care o avem astăzi sub un titlu mai poetic, „Insuportabilul automatism al ființei”:

> Cea mai mare parte a vieții psihice din fiecare clipă trebuie să apară prin mijloace nonconștiente, dacă e să apară [...]. Reglarea conștientă și intenționată a comportamentului, evaluărilor, deciziilor și stărilor noastre emoționale necesită un efort considerabil și este relativ lentă. Mai mult, ea pare că necesită o resursă limitată care se epuizează rapid, așa că actele de autoreglare conștientă nu pot avea loc decât în mică măsură și pentru scurt timp. Pe de altă parte, [procesele psihologice...] nonconștiente sau automate sunt neintenționate, facile, foarte rapide și multe dintre ele se pot desfășura în orice moment. Cel mai important, ele nu presupun efort, funcționează încontinuu, ghidându-l pe individ în siguranță pe parcursul zilei[8].

Totuși, recunoașterea modernă a ubicuității funcționării psihice inconștiente nu s-a bazat pe teoriile lui Freud. La aceasta s-a ajuns într-un mod cu totul independent, pe baza altor descoperiri neurologice și dovezi experimentale. Un punct de cotitură a fost cazul lui H.M., care a fost publicat în 1957. Ca urmare a rezecției chirurgicale a hipocampilor (regiunea cortexului responsabilă cu codificarea amintirilor declarative, cea care îi producea crize epileptice), era complet incapabil să-și amintească vreun eveniment petrecut după data intervenției. Cu toate acestea, neuropsihologii care l-au examinat în mod repetat de atunci (1953) până la moartea sa, în 2008, au observat că scorurile sale la testele psihometrice s-au îmbunătățit semnificativ de-a lungul anilor, demonstrând efecte clare ale exerciţiului. A învățat cum să se descurce la teste, deși nu își amintea în mod conștient că le mai făcuse[9].

Observații asemănătoare a făcut anterior şi Édouard Claparède, deși nu și-a dat seama de importanța lor. A raportat cazul unei femei amnezice care dădea mâna cu el în fiecare zi ca și cum nu îl mai întâlnise niciodată (ca pacientul meu, domnul S.). Claparède s-a hotărât să-și ascundă o piuneză în palmă, ca să vadă dacă ea va fi la fel de dornică să-l salute după ce a fost înțepată. După această experiență, pacienta a refuzat să dea mâna cu el, deși nu își amintea în mod conștient de evenimentul dureros. Când a întrebat-o de ce refuză, ea a dat doar explicații evazive, precum: „Oare doamnele nu au dreptul să nu dea mâna cu domnii?”[10].

O procesare inconștientă echivalentă a fost demonstrată și pentru apercepție. Înainte ca comisiile de control instituțional să-i evalueze etica protocoalelor de cercetare medicală, Roger Sperry a raportat cazul unei femei ale cărei emisfere cerebrale

fuseseră separate chirurgical pentru a controla o epilepsie netratabilă. Când emisfera sa dreaptă izolată a fost expusă la o succesiune rapidă de imagini pornografice, pacienta a chicotit și a roșit, deși emisfera ei stângă (verbală) nu a putut „să declare” în mod conștient despre ce sentimente era vorba. Ea a putut să facă doar niște afirmații tangențiale pentru a-și explica stânjeneala, precum „Ce mașinărie grozavă aveți, domnule Sperry!”[11].

Aici e important să observăm că procesele mintale inconștiente nu se limitează la structurile cerebrale subcorticale. În cazul lui Sperry, de exemplu, pacienta era incapabilă să conștientizeze informațiile perceptive stânjenitoare primite, procesate și recunoscute de cortexul ei cerebral drept.

Faptul că cortexul poate îndeplini asemenea funcții fără conștientizare a fost demonstrat în mod concludent de multe studii experimentale. Într-unul din ele, participanții la cercetare au fost expuși la o succesiune foarte rapidă de cuvinte pozitive și negative – atât de rapidă, încât nu au fost conștienți că au văzut ceva. Comportamentul lor ulterior a fost în mod clar influențat de cuvintele pe care pretindeau că nu le văzuseră: de exemplu, după expunerea la o succesiune rapidă de atribute negative asociate cu o fotografie a chipului A și de atribute pozitive asociate cu chipul B, participanții au manifestat o preferință pentru chipul B, deși nu știau de ce îl preferă[12]. Asta arată că cuvintele pozitive și cele negative trebuie să fi fost văzute, citite *și înțelese* în mod inconștient. Întrucât citirea implicând comprehensiunea este o funcție exclusiv corticală – exact de tipul celor considerate tipic „psihice” de anatomiștii clasici –, nu putem concluziona decât

că *funcțiile corticale nu sunt inerent conștiente*.

Aceasta este opinia general acceptată în știința cognitivă de astăzi: psihicul interpretat drept imagini mnezice (acum numite „reprezentări”) nu este conștient în mod intrinsec. Partea din creier care generează „conținutul” conștiinței („conștiința ca experiență”) – cortexul – poate să facă exact același lucru şi în lipsa experienței conștiente. Dar, dacă așa stau lucrurile, atunci ce anume face ca funcțiile sale inconștiente să fie *psihice*? Prin ce se deosebește cortexul – așa-zisul organ al minții – de alte dispozitive de procesare a informațiilor precum telefoanele mobile? Din nou, riscăm să devenim incoerenți.

Revenim astfel la marea întrebare cu care am încheiat capitolul 3: dacă nu din cortex vine conștiința noastră, atunci de unde? Și, mai mult, dacă cea mai mare parte a „vieții psihice din fiecare clipă” se desfășoară fără experiență conștientă, atunci de ce *uneori* presupune experiență conștientă? De ce toată această prelucrare a informațiilor nu se desfășoară nonconștient?

Istoria neuroștiinței cognitive este un cimitir al teoriilor care au încercat să specifice funcția conștiinței. Ele au în comun ipoteza că funcția sa este de „a lega” multitudinea fluxurilor de informație ce se întind prin tot creierul în întregul coerent care ne caracterizează experiența conștientă. De exemplu, cititul, recunoașterea facială, percepția culorilor, percepția mișcării, recunoașterea obiectelor, percepția spațiului și altele asemenea, toate acestea se întâmplă simultan în arii localizate în diverse părți ale creierului. Cum anume se îmbină ele în imaginile vizuale unificate pe care le percepem în mod obișnuit – cu culoarea și mișcarea unui chip, de exemplu, întâmplându-se

exact în locurile potrivite?

Meynert a afirmat că această funcție de conexare era îndeplinită de fibrele transcorticale, care „asociază” între ele imagini mnezice. Au trecut mai bine de o sută de ani, dar noi cam tot la această ipoteză am rămas. James Newman și Bernard Baars au sugerat că un „spațiu de lucru global” unificat este generat în cortex de talamus, făcând fragmentele disparate de informație accesibile în mod global experienței. Stanislas Dehaene și Lionel Naccache au adăugat că ariile corticale de asociere prefrontale și parietale integrează activitățile zonelor senzoriale primare în acest spațiu de lucru. Gerald Edelman a introdus „circuitele de reintrare” talamocorticale drept funcția esențială prin care informația integrată este trimisă înapoi la nivelurile anterioare ale procesării perceptive. Giulio Tononi a evidențiat procesul rezultant de prelucrare a informațiilor „integrate masiv”, afirmând că tocmai gradul de integrare a acestor fragmente este componenta-cheie; conștiința este o funcție a *cât de multe* informații se integrează. Francis Crick și Christof Koch au emis ipoteza că sincronia oscilațiilor gama din cortex conexează și stochează experiențele – cu alte cuvinte, că integrarea s-ar putea petrece în timp, și nu în spațiu. Rodolfo Llinás a sugerat în mod similar că activitatea talamocorticală se sincronizează sub 40 Hz. Și așa mai departe[13].

Niciuna dintre aceste teorii nu ne spune de ce sau în ce mod conexarea informațiilor – prin asociere, oscilare, sincronizare, reintrare, integrare masivă și altele asemenea – dă naștere *cu necesitate* experienței. De ce, mai precis, trebuie resimţit în mod conștient conținutul „spațiului de lucru global” pentru prelucrarea informațiilor? Oare conexarea, stocarea, sincronizarea, integrarea masivă și așa mai departe nu apar și

în cadrul prelucrării *inconștiente* a informațiilor? Calculatoarele generează spații de lucru globale și integrează masiv informații tot timpul, atunci când sunt interconectate prin internet. Și atunci de ce n-ar fi internetul conștient?

Un aspect îngrijorător, ce decurge direct din aceste teorii contemporane, este că tot mai mulți specialiști respectabili în neuroștiințe (precum Koch și Tononi) sugerează că acest lucru *ar fi* posibil. Asta din cauză că teoriile lor îi obligă să accepte această posibilitate ciudată. Procedând astfel, ei urmează curentul de gândire „panpsihist" inițiat de Thomas Nagel, conform căruia s-ar putea ca *toate* lucrurile să fie (puțin) conștiente[14].

Ar trebui să evaluăm aceste teorii în lumina observațiilor despre care am vorbit mai sus, în sensul că conștiința persistă în lipsa cortexului. Newman și Baars își localizează „spațiul de lucru global" în cortexul cerebral și totuși animalele supuse intervenției de decorticare cerebrală și copiii născuți fără cortex par să fie conștienți. Cel puțin, dovezile că aceste animale și aceşti copii sunt ființe conştiente sunt mult mai convingătoare decât dovezile că internetul ar fi conştient. Mai mult, adulții care pot să vorbească și al căror substrat așa-zis esențial pentru prelucrarea informațiilor integrate masiv a fost distrus de boală ne spun că simţul lor al sinelui persistă în lipsa substratului respectiv. Această afirmație este valabilă în cazul pacientului meu W., căruia îi lipseau complet lobii prefrontali.

Nimic din toate acestea nu ar trebui să ne surprindă dacă vom combina asemenea observații clinice cu dovezile experimentale analizate anterior, în sensul că cortexul efectuează *inconștient* cea mai mare parte a prelucrării informațiilor (cum sunt cititul și recunoașterea facială).

În dovezile pe care le-am prezentat există ceva remarcabil la care aș vrea să mă refer acum. Când pacienta lui Claparède a refuzat să dea mâna cu el, ea nu știa de ce. Totuși, probabil *a simțit* o anumită aversiune față de acest gest. Probabil avea vreo bază subiectivă când a refuzat să dea mâna cu el, deși nu avea acces la cauza obiectivă (amintirea piunezei). Același lucru e valabil și în cazul pacientei lui Sperry: s-a simțit stânjenită, dar nu știa de ce. Cu alte cuvinte, nu era conștientă de cauza obiectivă a sentimentului ei (imaginile pornografice), dar a conștientizat sentimentele subiective care le-au însoțit (stânjeneala).

Același lucru s-a întâmplat probabil și în experimentul cu succesiunea rapidă de cuvinte: participanții au simțit probabil o preferință pentru chipul B, deși le lipsea conștientizarea explicației acestui fapt. Același lucru pare să fie valabil și în cazul fetiței cu hidroanencefalie din figura 5: ea a manifestat o plăcere subiectivă când frățiorul său i-a fost așezat în poală, deși nu avea cum să știe cauza obiectivă a emoţiei sale, fiind complet lipsită de imagini corticale.

Prin urmare, în toate aceste cazuri, când am spus că pacienții și participanții la cercetare pe care i-am descris nu erau „conștienți” de cauzele acțiunilor lor, m-am exprimat prea vag. Ei nu erau conștienţi de anumite percepții sau amintiri – reprezentări –, dar *sentimentele* lor persistau. Exista încă „ceva cum e” să fii ei, care le determina judecățile de valoare. Erau conștienți de sentimentele lor, numai că nu le conștientizau sursa.

Acest lucru ne oferă un indiciu important cu privire la ce înseamnă în mod fundamental capacitatea de a simţi. Aparent

singurele printre funcțiile psihice, sentimentele sunt *cu necesitate* conștiente. Cine a auzit vreodată de un sentiment ce nu are o calitate subiectivă? Care ar fi rostul unui sentiment dacă nu l-ai simți? Chiar Freud a acceptat că sentimentele *trebuie să fie* conștiente:

> Din esența unui sentiment face totuși parte faptul că el e resimțit, că e cunoscut, așadar, de către conștiință. Posibilitatea unui inconștient nu s-ar pune absolut deloc atunci când vorbim de sentimente, senzații și afecte[*1] [15].

Îmi dau seama că unele persoane (inclusiv psihanaliști) resping această afirmație cu argumentul că există emoții inconștiente. Voi reveni curând la această chestiune. Deocamdată, aș vrea să fie cât se poate de clar cu ce sens folosesc termenul „sentiment”: mă refer la acel aspect al unei emoţii (sau al oricărui afect) pe care îl *simți*. Mă refer la ceea ce simţi ca atare. Dacă nu simți ceva, atunci nu vorbim de un sentiment.

Cu câteva paragrafe mai sus am spus că pacienta lui Claparède trebuie să fi simțit o oarecare aversiune, pacienta lui Sperry trebuie să fi simțit stânjeneală, participanții la testul cu succesiunea rapidă de cuvinte trebuie să fi simțit o preferință și așa mai departe. M-am exprimat astfel pentru că specialiștii în neuroștiințe de obicei nu pun întrebări despre experiențe subiective. Acesta este singurul motiv pentru care nu știm ce au simțit acești subiecți de cercetare. Însă eu (ca și Freud și Sacks) iau foarte în serios relatările introspective ale pacienților. În acest fel, sper să evit erori precum confundarea visului cu somnul REM.

Dacă o sută de oameni dintr-o sută raportează o senzație de durere când îi ciupești de mână, nu este irațional să concluzionezi că ciupirea mâinii cauzează durere, chiar dacă trebuie să te bazezi pe relatări introspective în acest sens în fiecare situație. Această afirmație este chiar mai valabilă pentru observațiile pe care le poți replica în cazul tău, când ai acces direct la fenomenul în cauză. Dacă mă doare când mă ciupesc de mână, atunci am o experiență personală a ceea ce vor să spună alții când raportează „durere”.

Sentimentele sunt dificil de studiat din cauză că sunt inerent subiective, dar (cu voia behavioriștilor) asta nu ne îndreptățește să le ignorăm. Dacă excludem sentimentele din descrierea creierului, nu vom înțelege niciodată cum funcționează acesta. Rolul jucat de sentimente în confabulațiile domnului S. ilustrează foarte bine această idee.

Când ne uităm la ce apare în conștiința noastră, observăm câteva categorii generale de conținut. Găsim, desigur, „reprezentările” lumii exterioare – atât percepții, amintiri ale ei, cât și gânduri despre ea. Filosofii acordă multă atenție acestor reprezentări, iar modelul empirist al minții a fost conceput pentru a le explica. Dar în conștiință mai găsim și altceva.

Găsim și sentimente – sentimente despre ce se întâmplă în lume, sentimente despre gândurile noastre în privinţa lumii, sentimente (mai ales) despre noi înșine, inclusiv sentimente ce par să fie relatări despre starea propriului corp. Există și sentimente vagi, sentimentele și dispoziţiile afective care descriu modul în care resimțim lumea și care ne modelează comportamentul în interiorul ei. Uneori, ele trec drept senzații

corporale, însă multe dispoziţii afective par a nu se datora nici stării corpului, nici vreunui alt lucru pe care îl putem identifica în lumea exterioară. Oare nu e conștiința plină de asemenea sentimente? Și totuși, e uimitor cât de frecvent sunt trecute cu vederea de specialiștii în neuroștiințe care caută o explicație a conștiinței[16].

E uluitor că filosofii empiriști și moștenitorii lor din domeniul științei, behavioriștii și cognitiviștii, au ignorat sentimentele[17]. Behavioriștii au afirmat că întreaga învățare este guvernată de „recompense” și „pedepse”, dar nu ne-au spus niciodată ce sunt acestea. Au realizat experimente riguroase care au generat „legea efectului”. Conform acesteia, dacă un comportament este urmat constant de recompense, se va intensifica, iar dacă e urmat constant de pedepse, se va diminua. Acest proces al învățării din experiență a fost numit „condiționare”.

Edward Thorndike, autorul legii efectului, a vrut să demonstreze că animalele învață prin încercare și eroare, nu judecând. Dar, de fapt, legea lui Thorndike echivalează cu o lege a *afectului*[18], deoarece sugerează că repetăm (și noi, și animalele) comportamentele care ne fac să ne simțim bine și că le evităm pe cele care ne fac să ne simțim rău. Așadar, legea efectului nu e în esența sa decât „principiul plăcerii” despre care vorbea Freud. Dar behavioriștii nu puteau să accepte existența a ceva atât de subiectiv precum sentimentele. E bine-cunoscut că B.F. Skinner, de exemplu, a declarat: „«Emoţiile» sunt exemple excelente ale cauzelor fictive pe seama cărora punem în mod obișnuit comportamentul”[19].

Felul cum şi-a formulat iniţial Thorndike legea dă în vileag minciuna: „Este mai probabil ca răspunsurile care produc un

efect satisfăcător într-o anumită situație să apară din nou în situația respectivă şi mai puţin probabil ca răspunsurile care produc un efect neplăcut să apară din nou în situația respectivă”[20]. Cuvintele „satisfăcător” și „neplăcut” au fost mai apoi eliminate și înlocuite cu „de întărire” și „punitiv”. Iată de ce:

> În psihologie, termenii noi, „de întărire” și „punitiv”, sunt folosiți altfel decât în sens colocvial. Existența a ceva ce întăreşte un comportament crește probabilitatea ca acel comportament să apară din nou, iar existența a ceva ce pedepsește un comportament reduce probabilitatea ca acel comportament să apară din nou[21].

Această definiție a termenilor e cu totul lipsită de conținut, după cum și trebuie să fie, întrucât cuvântul esențial este „ceva”. Aceeași afirmație e valabilă în cazul termenilor behavioriști „întărire pozitivă” și „întărire negativă”: ce anume face ca o întărire să fie pozitivă sau negativă dacă nu un sentiment?

Probabil se intenționa ca semnificația „recompensei” și „pedepsei” să fie că valoarea este intrinsecă *stimulului* – și nu receptorului stimulului. Dacă un cal se apropie de mine și eu îi dau un cub de zahăr, va exista (conform legii efectului) o probabilitate mai mare să se apropie de mine din nou, dar, dacă îl stropesc cu suc de lămâie în ochi, această probabilitate scade. Conform lui Thorndike, cubul de zahăr și lămâia devin *ele însele* în acest fel recompense sau pedepse ale comportamentului calului; nu e nevoie să luăm în calcul senzațiile pe care le evocă, dacă există așa ceva. Bineînțeles,

acesta este un raționament greșit, care plasează eronat forța cauzală, având drept consecință problema de care s-a plâns Sacks: un psihic lăsat fără intenționalitate. Una e să tratezi creierul ca pe o cutie neagră din rațiuni metodologice și cu totul altceva să atribui lucrurilor din afara cutiei puteri cauzale pe care de fapt ele nu le au, iar apoi să concluzionezi că în cutie nu se întâmplă nimic.

Ideea că sentimentele sunt „fictive” are multe efecte nocive asupra științei. De exemplu, în cea mai mare parte a secolului trecut, atunci când erau studiate mecanismele fiziologice de bază ale echilibrului energetic, behavioriștii au interzis folosirea cuvintelor „foame” și „sațietate”, întrucât asemenea lucruri nu pot fi văzute sau atinse. Behavioriștii vorbeau numai despre „stimulente” și „recompense” asociate cu comportamentul alimentar. Aceasta nu e doar o chestiune de semantică. Dacă un cuvânt precum „foame” nu poate fi folosit, atunci cum putem să înțelegem rolul pe care îl joacă ea în reglarea mâncatului? Oare acest lucru nu ar putea să întârzie, la rândul său, dezvoltarea unor tratamente pentru reducerea senzației de foame în cazurile de obezitate? De fapt, comportamentul alimentar este reglat de două mecanisme cerebrale ce interacționează între ele: un sistem „homeostatic”, ce reglează rezervele de energie, și un sistem „hedonist”, ce acționează asupra *apetitului*[22]. Și, la fel ca în cazul unor afecte corporale precum foamea, oare nu ar fi posibil ca interzicerea unor cuvinte denumind emoții precum „tristețe” și „frică” să întârzie dezvoltarea unor tratamente antidepresive și anxiolitice?[23] Dacă sentimentele chiar există, atunci cu siguranță că au corelative fiziologice care ne vor scăpa dacă nu le luăm în considerare.

În cele ce urmează mă voi despărți de behavioriști. Urmându-l pe Panksepp, primul specialist în neuroștiințe care a avut curajul să folosească cuvinte precum „foame” ca să explice reglarea echilibrului energetic atât la oameni, cât și la alte animale, voi accepta că sentimentele există *cu adevărat*[24]. Voi presupune că știi ce înseamnă să simți că îți este sete sau că ești trist, somnoros, amuzat, încrezător sau nesigur. Această ipoteză nu se justifică mai puțin decât alte inferențe științifice despre lucruri din natură. Și poate fi testată prin metodele obișnuite, prin predicții falsificabile.

Sentimentele sunt reale și știm de existența lor pentru că sunt peste tot în conștiința noastră. De fapt, ele sunt, din motivele pe care le voi explica în continuare, izvorul ființei conştiente – într-un sens care nu mi se pare deloc metaforic. De la originea lor în unele dintre cele mai străvechi straturi ale creierului, ele irigă solul sterp al reprezentărilor inconștiente și le aduc la viața psihică.

*1 Sigmund Freud, „Inconștientul”, traducere de G. Lepădatu, în *Opere esențiale*, vol III: *Psihologia inconștientului*. București: Trei, e-book.

5
Sentimentele

Spre încântarea medicilor lui și a noastră a tuturor, fratele meu Lee și-a revenit cu bine după leziunea la creier și s-a adaptat excelent. Deși putea fi indus în eroare cu ușurință de oameni care căutau să profite de el, din punct de vedere fizic era masiv și puternic și s-a folosit de acest lucru cu succes. Nu a mai fost niciodată exact același ca înainte de accident, dar și-a continuat viața.

Mie mi-a fost mai greu să mă adaptez.

Când aveam cinci ani, tata i-a cumpărat lui Lee un ceas de mână, pe care i-a promis că i-l dă când va învăța să spună cât e ora. A urmat o lecție scurtă. Lee s-a chinuit cu întrebări despre ce înseamnă când acul mic e la 9 și cel mare la 11. Observându-l din capătul opus al încăperii, am dat eu rapid răspunsurile. Tata mi-a spus să tac. În seara aceea i-a dat ceasul lui Lee, deși fratele meu încă nu se descurca bine. Și ceasul meu? mă gândeam.

Pe măsură ce am mai crescut, asemenea sentimente copilărești au făcut loc vinovăției. Eu mă descurcam bine la școală. Lee a fost norocos că a terminat-o. Pur și simplu, pentru mine lucrurile păreau să fie mai ușoare. Îmi doream ca familia

mea să poată vorbi despre această problemă, ca să găsească strategii care să mă ajute să nu îi creez lui neplăceri. Totuși, mama nu suporta să se discute despre accident, iar tata nu era genul de persoană cu care să pot discuta *orice* sincer – și în niciun caz despre cum îşi făcea el treaba de cap al familiei.

Astăzi pare ceva evident, dar, când m-am hotărât să studiez neuropsihologia, nu îmi trecea prin cap că ar putea să aibă vreo legătură cu accidentul lui Lee. Îmi amintesc însă că în copilărie mă gândeam că singurul lucru pe care merită să-l fac în viaţă este să aflu ce înseamnă „să exişti”. Abia în perioada în care am făcut studii de psihanaliză, după ani de zile, când deja lucram ca neuropsiholog, am reușit să pun totul cap la cap. În mod clar, cariera pentru care optasem era un compromis între ambiție, pe de o parte, și vinovăție, pe de altă parte. Dacă deveneam neuropsiholog, puteam să am succes din punct de vedere academic și, în același timp, să ajut oameni aflați în dificultate. De asemenea, probabil că asta explică de ce mă simțeam atât de frustrat de neuropsihologia pur academică și de ce m-am hotărât să devin clinician și, ocazional, chiar terapeut pentru neurorecuperare – roluri profesionale pe care specialiștii în neuroștiințe tind să le desconsidere.

Întrerup puțin șirul amintirilor. Ți se par toate acestea o explicație plauzibilă a deciziei mele? Aceasta nu este o întrebare-capcană: cred că amintirile ascunse ale acestor experiențe explică motivele alegerilor mele în carieră. Totuși, în general, oare nu așa par adesea că ne motivează sentimentele noastre, undeva sub pragul a ceea ce conștientizăm? E vorba de un loc comun al explicațiilor psihologice – un element de freudism atât de bine integrat în bunul-simț, încât puțini s-ar gândi să-l conteste. Păi, astfel, e ciudat să afirmi că sentimentele

sunt conștiente prin definiție: că sunt, într-un anumit sens, esența conștiinței. Cum să fie posibil așa ceva?

Într-o zi din 1985, după ce mi-am încheiat studiile universitare, m-am pornit spre spitalul Baragwanath pentru prima mea zi ca neuropsiholog stagiar. Contextul psihosomatic e că eram îngrijorat: profesorul meu – Michael Saling – își luase anul sabatic și plecase în Australia, eu rămânând fără supravegherea obișnuită. În plus, mă stresa și faptul că m-am rătăcit în drum spre spital, în Soweto, un complex urban foarte periculos pentru băieți albi ca mine.

Când, în cele din urmă, am găsit spitalul, ușurat că ajunsesem nevătămat, m-am repezit spre Secția 7 (neurochirurgie) și am întrebat de domnul Percy Miller, neurochirurgul consultant la care mi se spusese să mă prezint. Infirmiera-șefă m-a condus prin secție. Mirosul era puternic: un amestec de fluide corporale, antiseptic și ceva ce aducea a varză fiartă.

Miller semăna cu o pasăre. În timp ce stătea aplecat deasupra unui pacient căruia îi drena lichid cefalorahidian printr-un orificiu din cap, a început să-mi vorbească. „Ești Mark Solms? Încântat de cunoștință. Mike ne-a spus că-i ții locul. Mă bucur c-ai venit. Așteaptă puțin. Termin imediat și îți fac turul secției.”

Am încercat să par calm și profesionist. Să-i țin locul? Neurochirurgul şi-a spălat mâinile, apoi m-a dus la primul pat. Până să-mi dau seama ce se întâmplă, a început să turuie istoricul clinic al pacientului. „Acesta este domnul cutare, 43 de ani, astrocitom temporal stâng, grad III. Îl operăm miercuri. Evaluează-i în linii mari funcțiile limbajului și zi-ne și nouă. Și memoria, presupun; dumneata te pricepi mai bine decât mine.”

M-am gândit dacă să întreb ce e un astrocitom grad III.

Am trecut la patul următor. „Acesta este domnul cutare, 58 de ani. Are un macroadenom hipofizar. Îl operăm mâine, așa că te rog să-l evaluezi mai întâi pe el. Facem abordare transsfenoidală. Mike consideră că așa rezultatele cognitive sunt mai bune – ceea ce are sens.”

Și tot așa, de la un pat la altul. Când am ajuns la al cincilea sau al șaselea, era prea târziu să mărturisesc că nu înțelegeam mai nimic din ce mi se spunea. De asemenea, nu îmi puteam aminti detaliile cele mai importante ale primelor câteva cazuri. Abia dacă îmi aminteam câte ceva. Mi-am spus că voi reveni mai târziu ca să citesc fișele.

Am trecut prin niște uși batante și am intrat în salonul pentru femei. Chirurgul vorbea în continuare: „Aceasta este doamna cutare, 36 de ani, cisticercoză”. Nici asta nu știam ce înseamnă. Următoarea pacientă era complet goală. Și unii dintre bărbați erau goi, dar asta era altceva. M-am gândit că nu ar fi trebuit să văd așa ceva. Oare ei nu știu că nu sunt medic? Pe parcursul turului, observam expresiile faciale ale pacienților. Unul se uita la mine cu tristețe, altul absent, unii neliniștiți, mulți nu mă priveau deloc.

Eram complet depășit de situație. Oamenii ăștia presupuneau că știu ce fac. Și, pentru că nu am comentat această presupunere, era clar că urma să fiu responsabil pentru moartea cuiva.

În salonul de pediatrie a fost și mai rău. Primul copil era un caz grav de hidrocefalie: capul lui era cât al lui ET, de două ori mărimea normală. Mi s-a făcut greață, îmi era cald și transpiram abundent. Al doilea copil avea un balon moale de carne care îi ieșea pe la ceafă. Se uita țintă la perete. Percy

Miller a spus ceva despre un „mielomeningocel”. Intensitatea vocii lui s-a atenuat, înlocuită de bătăile lente ale inimii mele. Mi s-a îngustat câmpul vizual și, în timp ce începeam să nu îi mai văd gura, care vorbea, am avut un gând: o să mă lovesc cu capul de podea, o să-mi sparg capul și-o să ajung într-unul din paturile astea. M-am simțit ciudat de ușurat, apoi mi-am pierdut cunoștința.

Acest exemplu al primei mele zile la serviciu ilustrează multe dintre caracteristicile esențiale ale „afectului”, inclusiv faptul că, deși nu știm întotdeauna de ce simțim ce simțim, cu siguranță știm *ce* simțim. În acest caz, am simțit că urma o catastrofă, pe care am legat-o din punct de vedere intelectual de incompetența mea profesională, în spatele căreia se aflau tot felul de sentimente complicate referitoare la fratele meu, incluzând atât nevoia mea de a-l ajuta, cât și o identificare cu el la un nivel profund. Dar cauza imediată a leșinului meu a fost un reflex *corporal* primitiv, declanșat direct de emoţiile mele.

Sincopa vasovagală te face să leșini deoarece creierul reacționează la ceva alarmant, de obicei la vederea sângelui sau la alt risc perceput de accidentare fizică. Acest declanșator (înregistrat de amigdală) activează nucleul solitar din trunchiul cerebral, ceea ce duce la scăderea bruscă a pulsului și a tensiunii arteriale. Iar asta duce la un aport sangvin redus în creier; și îți pierzi cunoștința.

De ce avem acest reflex înnăscut? Răspunsul este: el reduce fluxul sangvin și, astfel, oprește o hemoragie, în anticiparea unei leziuni[1]. Acest reflex duce la leșin numai la oameni, din cauza staţiunii noastre verticale și a creierului mare, ce necesită un efort cardiac mai mare. (Prin urmare, e puțin

probabilă teoria că leșinul este o formă de „a face pe mortul”.) În linii mari, reflexul vasovagal crește șansele să supraviețuiești unei accidentări, nu le reduce; și, în orice caz, fluxul sangvin cerebral se restabilește imediat ce cazi la pământ.

Să luăm alt exemplu[2]. Controlul respirator este în mod normal automat: cât timp nivelurile de oxigen și dioxid de carbon din sânge rămân în limite viabile, nu e nevoie să conștientizezi respirația ca să respiri. Totuși, atunci când gazele sangvine depășesc aceste limite normale, controlul respirator intră peste conștiință sub forma unei senzații acute numite „foame de aer”. Valorile neprevăzute ale gazelor sangvine sunt un semnal că *trebuie luate măsuri*. E o nevoie urgentă să eliberezi o obstrucție a căilor respiratorii sau să ieși dintr-o încăpere plină de dioxid de carbon. În acest moment, controlul respirator este conștientizat printr-un sistem intern de avertizare pe care îl resimțim ca pe o *alarmă* – în mod specific, în acest caz, alarmă de sufocare.

Unele dintre cele mai simple forme de senzații – foamea, setea, somnolența, oboseala musculară, greața, frigul, urgența urinară, nevoia de a defeca și altele asemenea – poate că nu par afecte, dar ele asta sunt. Stările afective se deosebesc de alte stări psihice prin faptul că au *valențe* hedonice: se simt ca „bune” sau „rele”. În asta constă diferența dintre senzațiile afective precum foamea și setea și cele senzoriale precum văzul și auzul. Imaginea și sunetul nu au o valoare intrinsecă, spre deosebire de senzațiile afective[3].

„Binele” sau „răul” unei senzații îți spune ceva despre starea nevoii biologice din spatele ei. Astfel, setea e ceva rău, iar potolirea ei e ceva bun, întrucât e necesar să-ți menții nivelul de hidratare în limitele viabile pentru supraviețuire. Aceeași

afirmație e valabilă în cazul senzației neplăcute de foame în relaţie cu ușurarea plăcută pe care ți-o dă mâncatul. Pe scurt, plăcerea și neplăcerea îți spun *cum te descurci* în legătură cu nevoile tale biologice. Valența reflectă sistemul de valori ce susține toate aspectele vieții biologice, şi anume că e „bine” să supraviețuiești și să te reproduci și „rău” să nu faci asta.

Desigur, individul nu e motivat direct de aceste valori biologice, ci de senzațiile subiective generate de ele – chiar dacă nu avem nicio idee care sunt valorile biologice de bază și chiar dacă nu le susținem din punct de vedere intelectual. De exemplu, mâncăm alimente dulci pentru că au gust bun, și nu pentru că tind să aibă un conținut caloric ridicat, care este motivul biologic pentru care au gust bun. Afectele spun povești lungi despre evoluție pe care nu le conștientizăm deloc. Ca în cazul reflexului vasovagal, conștientizăm numai senzațiile.

Motivul pentru care folosesc cuvântul „neplăcere”, și nu „durere” este că în creier există multe tipuri diferite de plăcere și de neplăcere. Foamea e o senzație neplăcută și te simți bine când o atenuezi mâncând; intestinele balonate îți dau o senzație neplăcută și te simți bine când o atenuezi defecând; durerea e o senzație neplăcută și te simți bine dacă te îndepărtezi de sursa ei. Acestea sunt afecte corporale, dar același lucru se aplică și celor emoționale. Detresa cauzată de separare e neplăcută și răspundem prin a căuta reîntâlnirea. Frica e o senzație neplăcută și scăpăm de ea îndepărtându-ne de pericol (și, uneori, leșinând). Alarma de sufocare, foamea, somnolența și frica sunt senzații neplăcute, dar din puncte de vedere diferite. Dar, dacă scăpăm de ele, ne simţim bine, tot din puncte de vedere diferite.

Senzațiile diferite semnalează situații diferite semnificative

din punct de vedere biologic și fiecare dintre ele ne obligă *să facem* ceva diferit: urinarea nu poate satisface foamea, iar mâncatul nu poate ușura o vezică urinară plină. Să ne amintim ce a spus Damasio despre pacientul B.: „Raporta şi senzații de foame, de sete și dorința de a excreta *și avea comportamentele aferente*”. Ar fi fatal dacă am proceda altfel.

Senzațiile ne determină să facem ceva *necesar*. În acest sens, ele sunt indicatori ai cererilor de travaliu. În cazul foamei de aer, travaliul necesar trebuie să-ți reechilibreze gazele sangvine. În cazul hipotermiei, travaliul trebuie să te readucă într-un interval de temperatură viabil. În cazul detresei cauzate de separare, travaliul trebuie să te ducă la reîntâlnirea cu cineva care îți poartă de grijă. Și așa mai departe. În jargonul teoriei controlului, dezechilibrele gazelor sangvine, valorile suboptime ale temperaturii, lipsa celor care îți poartă de grijă și prădătorii ce se apropie sunt „semnale de eroare”, iar acțiunile pe care le generează sunt menite *să corecteze* erorile. Rezolvarea unui afect prin ceva de tipul satisfacerii înseamnă că o eroare a fost corectată cu succes, după care dispare de pe radarul conștiinței.

Am ajuns din nou pe un teritoriu ce li se va părea ciudat de familiar psihanaliștilor. Ne amintim că Freud a definit „pulsiunea” ca „o măsură a cererii de travaliu ce e impusă psihicului ca urmare a legăturii sale cu ceea ce este corporal”. Acum se vede cât de strâns legate sunt afectele de pulsiuni – ele sunt manifestările lor subiective. Afectele sunt modul în care devenim conștienți de pulsiunile noastre, ele ne spun cât de bine sau cât de rău merg lucrurile în legătură cu nevoile specifice pe care le evaluează.

Iată pentru ce sunt afectele: ele transmit *ce* aspecte biologice

merg bine sau rău pentru noi și ne impulsionează să luăm măsuri. În acest sens, senzațiile afective sunt diferite de cele perceptive. Filosofii sunt adesea neliniștiți de posibilitatea a ceea ce ei numesc „qualia inversate”. Cum pot să știu că roșul pe care îl văd eu este la fel ca roșul pe care îl vezi și tu? Oare nu e posibil ca roșul meu să ți se pară albastru? Problema altor minți sugerează că nu putem ști niciodată, pentru că și tu, și eu am indica aceleași obiecte din lume și le-am numi „roșii”. Dar ce e valabil despre percepțiile vizuale nu e valabil despre experiența afectivă. Calitatea de a fi roșu nu *cauzează* nimic diferit de calitatea de a fi albastru, astfel încât le poți schimba între ele în mod arbitrar fără consecințe fizice.

Acest lucru nu e valabil în cazul unor senzații precum foamea în relaţie cu urgența urinară sau precum frica în relaţie cu detresa cauzată de separare. O senzație (frica) te constrânge să scapi de ceva; alta (detresa de separare) te constrânge să cauți pe cineva. Senzația nu poate fi separată de starea corporală pe care o aduce cu sine[4]. Dacă le-ai schimba între ele, ai simți o dorință irezistibilă de a scăpa de cineva care îți poartă de grijă, dar lipsește și ai căuta plângând un prădător care te tot urmărește. Dacă ai schimba între ele calitatea subiectivă de a fi roșu și cea de a fi albastru, nu ar exista consecințe, dar schimbarea senzației de frică cu detresa de separare (sau foamea cu urgența urinară) te-ar ucide[5].

Al doilea lucru de remarcat despre senzații e că *sunt întotdeauna conștiente.* O senzație de care nu ești conștient nu este o senzație. Acest lucru e adevărat prin definiție, după cum am spus mai sus, dar și datorită unei trăsături specifice a fiziologiei creierului. Vom afla de ce stau lucrurile așa în capitolul următor, unde discutăm despre mecanismele

cerebrale de „activare”. Deocamdată, vreau să te conving că senzațiile sunt întotdeauna conștiente, fără excepție. Nu vreau să spun că toate mecanismele de reglare a nevoilor din creier sunt conștiente, ci că *contează dacă o nevoie este simțită sau nu*. E posibil ca undeva, în fundal, proporția apă versus sare din corp să fluctueze tot timpul, dar, când o simți, vrei să bei. E posibil să fii în pericol în mod obiectiv fără să-ți dai seama, dar, când simți asta, cauți modalități de scăpare.

Pentru lucruri diferite e nevoie de nume diferite, iar diferența dintre nevoile resimțite și cele neresimțite face necesară introducerea unei distincții terminologice. „Nevoile” sunt diferite de „afecte”. Nevoile corporale pot fi înregistrate și reglate *autonom*, ca în exemplele controlului cardiovascular și respirator, termoreglării și metabolizării glucozei. Acestea se numesc funcții „vegetative”, și pe bună dreptate: nu au nimic conștient. De unde și termenul „reflex” autonom. Conștiința intră în ecuație numai atunci când nevoile sunt simțite – momentul în care *îți* solicită un travaliu. (Notă: pulsiunile evaluează solicitările de travaliu adresate *psihicului;* unele nevoi nu duc niciodată la acțiune voluntară și, de aceea, nu devin niciodată conștiente. Tensiunea arterială este un exemplu notoriu din punct de vedere clinic; nu știi nimic despre valorile ei prea mici sau prea mari până ce este deja prea târziu.)

Și nevoile emoționale pot fi gestionate automat, prin stereotipuri comportamentale precum „instinctele” (strategii de supraviețuire și de reproducere înnăscute, pe care Freud le-a plasat în centrul conceptului său de psihic inconștient). Dar nevoile emoționale sunt de obicei mai dificil de satisfăcut decât cele corporale, din rațiuni pe care le voi explica în capitolul de față. Acesta este motivul pentru care senzațiile pe care le evocă

durează de regulă mai mult. O senzație dispare din conștiință atunci când nevoia pe care o anunță a fost satisfăcută.

Al treilea lucru de remarcat în legătură cu senzațiile este că nevoile resimțite sunt *prioritizate* în detrimentul celor neresimțite. Suntem permanent asaltați de numeroase nevoi. Funcțiile vegetative precum echilibrul energetic, controlul respirator, digestia, termoreglarea și altele asemenea acționează constant, la fel și comportamentele stereotipe de tipuri diferite. Nu ai putea simți toate aceste lucruri în același timp, în special pentru că poți *face* câte un singur lucru (sau foarte puține lucruri) pe rând. Trebuie făcută o selecție. Ea are loc în funcție de *context*. Prioritățile se stabilesc în funcție de intensitatea *relativă* a nevoilor (amploarea semnalelor de eroare) *în raport cu* oportunitățile existente în circumstanțele curente. Iată un exemplu simplu. În timp ce țin o prelegere, vezica urinară mi se umple treptat, dar nu simt presiunea tot mai mare până ce îmi închei discursul, moment în care simt brusc nevoia să urinez. *Devin conștient* de semnalul de eroare datorită schimbării contextuale. În acel moment, simțind ocazia, nevoia mea corporală devine conștientă sub formă de senzație[6].

Prioritizarea nevoilor în acest fel are consecințe importante. Cea mai importantă este că, *atunci când devii conștient de o nevoie, când o resimți, ea îți domină comportamentul voluntar*.

Ce înseamnă „voluntar”? E opusul lui „automat”. Înseamnă supus unor *alegeri* ce țin de aici-și-acum. Alegerile pot fi făcute numai dacă se bazează pe un sistem de valori – ceea ce determină „calitatea de a fi bun” versus „calitatea de a fi rău”. Altfel, răspunsurile noastre la evenimente nefamiliare ar fi aleatorii. Iar aici închidem cercul, revenind la trăsătura

fundamentală a afectului: valența. Tu decizi ce faci și ce nu faci în baza *consecințelor resimțite în urma acțiunilor tale*. Aceasta este legea afectului. Astfel, comportamentul voluntar, ghidat de afect, are un enorm avantaj adaptativ față de comportamentul involuntar: ne eliberează de lanțurile automatismului și ne permite să supraviețuim în situații neprevăzute.

Faptul că comportamentul voluntar trebuie să fie conștient relevă cea mai profundă funcție biologică a senzației: ne ghidează comportamentul în condiții de *incertitudine*. Ne permite să determinăm în toiul acțiunii dacă un mod de a acționa este mai bun sau mai prost decât altul. În exemplul foamei de aer, reglarea gazelor sangvine devine conștientă atunci când nu ai o soluție de-a gata ca să te menții în limitele tale viabile din punct de vedere fiziologic. În graba de a scăpa din încăperea plină cu dioxid de carbon, de pildă, cum știi încotro s-o apuci? Nu ai mai fost niciodată în această situație (în nicio clădire în flăcări, ca să nu mai vorbim de exact această clădire), așa că nu ai cum să anticipezi ce e de făcut. Acum trebuie să te hotărăști dacă o iei pe aici sau pe acolo, în sus sau în jos etc. Iei asemenea decizii *în funcție de ceea ce simți în timp ce cauţi o soluţie la problemă*: senzația de sufocare se intensifică sau se diminuează, în funcție de direcția în care te îndrepți, cea greşită sau cea bună – adică în funcție de creșterea sau de scăderea nivelului de oxigen din aer.

Senzația conștientă a alarmei de sufocare implică diverse structuri neuronale și substanțe chimice dintre cele responsabile pentru controlul respirator inconștient, la fel cum senzația de foame selectează diferite sisteme cerebrale dintre cele responsabile cu reglarea autonomă a echilibrului energetic. Știința nu ar fi descoperit niciodată aceste lucruri

dacă ar fi continuat să ignore senzațiile.

Sunt mult mai multe de spus despre senzaţii, cum ar fi modul în care facilitează *învățarea din experiență* (prin legea afectului) și modul în care sunt legate de gândire, dar ideile pe care tocmai le-am exprimat, pe scurt, explică *de ce* simțim. Am descris ce adaugă ceea ce simţim la repertoriul mecanismelor pe care noi, ființele vii, le folosim pentru a rămâne în viață și a ne reproduce. Aceasta este contribuția psihologiei la biologie. Selecția naturală a influențat aceste mecanisme de supraviețuire, dar, odată ce s-au dezvoltat senzaţiile afective – adică abilitatea noastră unică de organisme complexe de *a lua notă de propriile stări* –, în univers a apărut ceva cu totul nou: ființa subiectivă.

E greu de imaginat cum ai putea să ai o durere fizică fără să o simți: să ai o durere pur și simplu înseamnă să o simți. Dar ce putem spune despre stările afective mai subtile pe care le numim emoții și dispoziții? Trebuie să simți că ești fericit ca să fii fericit? Oare nu e ceva obișnuit să realizezi prea târziu că ești stăpânit de o dispoziție proastă – să ajungi la concluzia că probabil ești deprimat și așa mai departe?

Am început să explorăm afectul cu ajutorul formelor sale corporale, pentru că acestea sunt exemplele cele mai simple și fără îndoială că au fost și primele care au apărut în cursul evoluției. Consider că „zorii conștiinței” nu au presupus ceva mai elaborat decât senzații somatice cu valenţe hedonice. Acum vreau să arăt că emoțiile umane sunt versiuni complexe ale unor asemenea lucruri. Și ele sunt, în cele din urmă, semnale „de eroare” ce iau notă de devieri de la stările tale biologice preferate și îți spun dacă măsurile pe care le iei îți

îmbunătățesc sau îți înrăutățesc situația.

Din păcate, nu există o clasificare a afectelor general acceptată în neuropsihologia curentă. Am făcut o distincție între afectele corporale și cele emoționale, dar asemenea delimitări precise nu există în natură. Fac această delimitare pornind de la taxonomia lui Jaak Panksepp, care este larg – dar nu unanim – acceptată. El a împărțit și mai departe afectele corporale, diversitatea lor fiind foarte mare, în subtipurile interoceptive („homeostatice") și exteroceptive („senzoriale"). Foamea și setea, de exemplu, sunt afecte homeostatice, iar durerea, surpriza și dezgustul sunt afecte senzoriale[*1]. Așadar, pentru a fi clar, conform lui Panksepp există trei tipuri de afecte: homeostatice și senzoriale (ambele tipuri fiind corporale) și emoționale (care implică trupul, dar nu pot fi descrise drept „corporale" în niciun sens simplu). Gândește-te, de exemplu, că ți-e dor de fratele tău, iar asta e o stare emoțională; nu este corporală în același fel în care sunt foamea și durerea.

Panksepp și-a fundamentat taxonomia pe studii asupra *stimulării cerebrale profunde* pe care el, studenții lui și mulți alții înaintea lor le-au efectuat pe mii de animale. I-am vizitat de multe ori laboratorul, care era o veritabilă menajerie, uneori plin de porumbei, alteori plin de pui de găină, câini, cobai, șobolani sau șoareci-de-câmp. (E posibil ca boala care l-a ucis în cele din urmă să fi fost cauzată de expunerea în exces la unele dintre aceste animale, mai ales la păsări.)

Nu încape îndoială că Jaak a simțit compasiune, chiar dragoste față de aceste animale. Dar trebuie să recunoaștem și că au fost sacrificate pe altarul științei, în număr mare, fără ca ele să-și aleagă soarta. E o ironie tristă că datorăm

experimentelor lui Panksepp informația aproape sigură că animalele enumerate mai sus sunt creaturi capabile să simtă, cu emoții intense, care, în esență, nu sunt diferite de cele resimțite de oameni. Ca urmare a acestor descoperiri și a preocupărilor tot mai mari pentru problemele de etică ce decurg din ele, Panksepp și-a petrecut ultimele decenii de viață studiind numai emoțiile pozitive.

În cele ce urmează, când trec de la observații despre animale la observații despre oameni și invers, o fac în mod deliberat. Când colegii l-au acuzat de antropomorfism față de animale, Panksepp a spus că mai degrabă s-ar recunoaște vinovat de zoomorfism față de oameni. Obiectivul experimentelor sale a fost să identifice structurile și circuitele cerebrale care activează în mod cert aceleași răspunsuri afective, nu numai la indivizi, ci și la mai multe specii. Când e vorba de afecte emoționale, s-a dovedit că șapte dintre ele pot fi reproduse nu numai la toate primatele, ci chiar *la toate mamiferele*, prin stimularea exact acelorași elemente chimice și structuri cerebrale. (Multe dintre ele pot fi evocate și la păsări, iar unele la toate vertebratele.) Mamiferele s-au separat de păsări acum aproximativ 200 de milioane de ani, iată cât de vechi sunt aceste emoții. Totuși, întrucât oamenii sunt mamifere, în continuare mă voi axa pe aceste șapte tipuri. Din câte știm, ele sunt ingredientele *de bază* ale întregului repertoriu emoțional uman. Numeroasele noastre bucurii și tristeți par să fie outputurile acestor șapte sisteme, amestecându-se între ele și cu procese cognitive superioare.

Probabil cea mai cunoscută taxonomie alternativă a „emoțiilor de bază” (așa cum sunt numite) este cea elaborată de Paul Ekman[7]. Diferențele apar în principal din faptul că Ekman

a folosit o metodă diferită de cea a lui Panksepp, şi anume studiul expresiilor faciale și al comportamentelor asociate. După cum a observat Charles Darwin cu mult timp în urmă în *Expresia emoțiilor la om și animale* (1872), acestea au elemente comune remarcabile la diverse specii de mamifere. Totuși, deși Panksepp și Ekman clasifică afectele în mod diferit (de exemplu, Ekman consideră că dezgustul este un afect „emoțional”, iar Panksepp îl consideră „senzorial”), există un larg acord în privința afectelor. Nimeni nu se îndoiește că dezgustul există, așa că, într-un anumit sens, nu prea contează cum îl clasificăm.

Principala opinie diferită este cea a Lisei Feldman Barrett. Din nou, dezacordul poate fi pus în primul rând pe seama unor diferențe metodologice[8]. Ea se axează pe emoțiile autoraportate la oameni și, deloc surprinzător, descoperă că există variații considerabile în ce privește modul în care diverși oameni (și diverse culturi) caracterizează și analizează sentimentele. Asta nu contrazice faptul că tipurile naturale de bază se află sub suprafața construită social. Voi ilustra imediat mecanismele prin care apare această variabilitate, dar explicația pe scurt este: reflexele și instinctele sunt instrumente nerafinate, dar funcționale pentru supraviețuire și succes reproductiv, dar ele nu ne pot echipa adecvat pentru multitudinea de situații imprevizibile și medii în care ajungem. Prin urmare, trebuie să completăm în mod adaptativ răspunsurile înnăscute cu învățarea din experiență. Faptul că oamenii fac acest lucru cu atâta ușurință e principalul motiv pentru care, până la urmă, am ajuns să dominăm lumea într-o asemenea măsură.

Programele instinctive aflate la baza acțiunilor oamenilor sunt de regulă atât de condiționate prin învățare, încât nu mai

pot fi recunoscute drept „instinctive”. Totuși, instinctele și reflexele sunt întotdeauna în fundal. Întreaga teorie psihanalitică se întemeiază pe ideea că, dacă îți bați capul să le găsești, tendințele instinctive implicite pot fi identificate întotdeauna în spatele intențiilor explicite.

Iar acum, o prezentare rapidă a emoțiilor instinctive, conform taxonomiei lui Panksepp[9]. Diverși oameni de știință clasifică afectele în mod diferit, așa că folosesc și cuvinte diferite. Panksepp a scris cu majuscule denumirile emoțiilor de bază pentru a le deosebi de utilizarea colocvială – adică pentru a arăta că vorbește despre funcțiile cerebrale sub toate aspectele lor, nu numai despre sentimente.

(1) DORINȚA SEXUALĂ. Nu suntem permanent excitați sexual. Sentimentul erotic intră în conștiință numai atunci când sexul este prioritizat în fața altor motivaţii, iar asta se întâmplă în contextul unor nevoi și oportunități fluctuante. Când dorința sexuală este activată, o simți, apoi sentimentele erotice îți dirijează acțiunile. Când ești excitat sexual, acorzi atenție altor detalii decât atunci când îți este frică, de exemplu, și te și comporți diferit. Astfel, conștiința exteroceptivă și comportamentul tău voluntar sunt determinate de starea ta interioară; resimți lumea în mod diferit – literalmente aducându-ți pe cap experiențe diferite –, în funcție de ceea ce simți. Acesta este și motivul pentru care nu poți simți în același timp excitare sexuală și repulsie motivată de frică, nu le poți prioritiza pe amândouă. Când nevoia de siguranță este prioritizată, mintea îți anulează motivaţiile sexuale.

Nu e clar dacă DORINȚA SEXUALĂ ar trebui clasificată ca

afect „corporal” sau „emoțional”. Unii se îndoiesc chiar că sexualitatea este o nevoie. Acesta este un exemplu excelent al diferenței dintre nevoi (inconștiente) și afectele (conștiente) pe care le generează. Când ne implicăm în acte sexuale, de obicei nu încercăm să ne îndeplinim datoria biologică. De fapt, foarte frecvent, sperăm să *nu* ne reproducem. Ca în cazul gustului dulce versus rezervele de energie, ceea ce ne motivează ca ființe subiective este căutarea plăcerii erotice, și nu a succesului reproductiv. Adică suntem conduși de *senzații*. Dar organismele vii *au nevoie* să se reproducă, cel puțin în general. Iată principalul motiv pentru care sexul a devenit plăcut din punct de vedere subiectiv, prin selecție naturală.

Am spus „în general” deoarece nu întreaga activitate sexuală duce la reproducere, ci numai suficient din ea cât să existe specia. Acest lucru ilustrează alt principiu central: utilitatea limitată a comportamentelor înnăscute pentru a ne satisface nevoile emoționale. În actul sexual, aspectele înnăscute se reduc doar la congestie și lubrifiere genitală, lordoză (arcuirea spatelui, care face vaginul disponibil pentru penetrare), încălecare, intromisiune, împingere și ejaculare. Împreună cu aceste reflexe, atingerea clitorisului sau a penisului (care sunt organe echivalente) într-un anumit ritm produce senzații plăcute care anticipează eliberarea tensiunii sexuale prin orgasm, în cele din urmă, până la satisfacere. Aceste mecanisme involuntare nu ne echipează pentru misiunea dificilă de a-i convinge pe alți oameni – mai ales pe cei de care ne simțim atrași – să se supună dorinței noastre de a face sex cu ei. Motivul principal pentru care nevoile „emoționale” sunt mai dificil de satisfăcut decât cele corporale este că, de regulă, ele implică alți agenți capabili de a simţi, care au propriile nevoi;

ele nu sunt simple substanțe precum mâncarea și apa. Prin urmare, pentru a ne satisface nevoile sexuale, trebuie să ne suplimentăm cunoașterea înnăscută cu alte abilități, dobândite prin *învățare*. Numai astfel se explică marea diversitate a activităților noastre sexuale, alături de forma „medie" moștenită prin selecție naturală.

Să observăm că învățarea nu *șterge* reflexele și instinctele; le elaborează, le completează și le domină, dar ele există în continuare. Felinarele stradale iluminează drumurile noaptea, dar nu pot să elimine complet întunericul. Mecanismul obișnuit de actualizare a memoriei pe termen lung, „reconsolidarea" (pe care îl descriu în capitolul 10), nu se aplică în cazul reflexelor și instinctelor. Iar asta pentru că reflexele și instinctele nu sunt *amintiri*, ci înclinații fundamentale „determinate genetic" în fiecare specie prin selecție naturală.

Varietatea comportamentelor noastre sexuale se lărgește și mai mult datorită faptului că circuitul cerebral al DORINȚEI SEXUALE tipice indivizilor de sex feminin și al celei tipice indivizilor de sex masculin există la fiecare mamifer. Tendința care ajunge să domine este determinată de factori diverși, inclusiv de evenimente genetice și intrauterine[10]. Nu voi intra aici în detalii de anatomie și de chimie, subliniez doar că ambele circuite apar în hipotalamus și se termină în substanța cenușie periapeductală, abreviată PAG. (Am arătat localizarea PAG în figura 6. Aflăm imediat de ce este atât de importantă.) Cu alte cuvinte, ele sunt în întregime subcorticale[11].

(2) CĂUTAREA. Toate nevoile corporale (și cele sexuale) – care sunt înregistrate de „detectoarele de nevoi" localizate în principal în hipotalamusul medial – activează această a doua

pulsiune emoțională, pe care am prezentat-o în capitolul 1. Este aproape sinonimă cu conceptul de „libido” al lui Freud, dar Freud nu știa că DORINȚA SEXUALĂ doar *activează* acest sistem; DORINȚA SEXUALĂ și CĂUTAREA nu sunt același lucru. CĂUTAREA generează un comportament explorator de „căutare a hranei”, însoțit de o stare emoțională conștientă ce poate fi caracterizată drept expectație, interes, curiozitate, entuziasm sau optimism. Să ne gândim la un câine în aer liber: indiferent care îi sunt nevoile corporale curente, căutarea hranei îl impulsionează să interacționeze pozitiv cu mediul, astfel încât să și le poată satisface. Aproape tot ce avem nevoie ca ființe vii se găsește „undeva afară”. Prin căutarea hranei învățăm, aproape accidental, ce lucruri din lume ne satisfac fiecare dintre dorințe. Astfel, codificăm relațiile lor de tip cauză-efect. Acest lucru ilustrează din nou modul în care instinctele stereotipe duc la învățare individualizată.

CĂUTAREA este neobișnuită printre emoțiile de bază prin faptul că acceptă incertitudinea *în mod proactiv*. Aceasta este originea comportamentelor de căutare a noutății și chiar de asumare a riscului. Căutarea hranei ne face să explorăm lucruri interesante, astfel încât să știm la ce să ne așteptăm când le vom mai întâlni. De exemplu, odată ce un câine a explorat un gard viu și s-a familiarizat cu conținutul lui, va fi mai puțin interesat de el data viitoare. În consecință, CĂUTAREA este emoția noastră „implicită”. Atunci când nu suntem dominați de vreunul dintre celelalte afecte („legate de sarcini”), conștiința noastră tinde spre acest sens generalizat al interesului față de lume.

Din punct de vedere anatomic, neuronii circuitului CĂUTĂRII apar din aria tegmentală ventrală a trunchiului

cerebral, de unde urmează un curs ascendent, prin hipotalamusul lateral spre *nucleus accumbens*, amigdală și cortexul frontal (vezi figura 2). Din punct de vedere chimic, neuromodulatorul care îl comandă este dopamina, „materia viselor” (vezi capitolul 1)[12]. Asta ne spune ceva interesant despre CĂUTARE, şi anume că poată fi activată chiar și în timpul somnului, prin solicitări de travaliu adresate psihicului, ducând la activități de rezolvare a problemelor ce trebuie ghidate de sentimente conștiente. De aici, visele.

(3) FURIA. Deși abordăm mediul într-un mod pozitiv prin CĂUTARE, cu credința optimistă că ne va satisface nevoile, lucrurile nu merg întotdeauna bine pentru noi. Așa cum preistoria evoluției ne-a dotat cu reflexe și instincte ce anticipează destul de bine căile pe care ne putem satisface nevoile corporale, tot astfel ne naștem cu tendințe emoționale ce anticipează căi pe care putem ieși din încurcătură. În situații complicate cu semnificație universală, sentimentele adecvate sunt prioritizate pentru a domina comportamentul. Astfel suntem scutiți de costurile biologice ale necesităţii de a reinventa roțile care le-au permis strămoșilor noștri să supraviețuiască și să se reproducă. Emoțiile sunt o moștenire prețioasă. Ele transmit abilități de supraviețuire înnăscute – cunoaștere inconștientă, implicită – sub forma conștientă a senzațiilor care ne pot ghida acțiunile în mod explicit.

Când se declanșează sistemul FURIEI – așa cum se întâmplă când intervine ceva între noi și orice ne-ar putea satisface nevoile curente –, conștiința este definită de sentimente ce variază de la frustrare iritată la furie oarbă. Reflexele și instinctele ce se manifestă în asemenea situații includ

piloerecția (erecția părului), protruzia ghearelor, sâsâiturile, mârâielile și arătarea colților, urmate de „atacul afectiv”: te arunci asupra țintei mâniei tale și o muști, o împingi sau o lovești până când se potolește.

De ce *simți* afectele care însoțesc asemenea comportamente? Răspunsul este ca și mai sus: sentimentele îți spun cum te descurci, dacă lucrurile merg bine sau prost, când încerci să scapi de un obstacol – care, adesea, încearcă să scape de tine în același timp. Simți gustul dulce al victoriei sau amarul înfrângerii. Acest lucru ghidează ce faci în continuare, inclusiv posibilitatea ca durerea (un afect senzorial suprimat în timpul atacului afectiv) să fie prioritizată, înlocuind astfel FURIA și punând capăt luptei – și poate ducând la fugă.

Cum ar putea toate acestea să se desfășoare automat, fără o evaluare conștientă continuă? Întrebarea se aplică cu atât mai mult rolului afectului în *gândire* – subiect pe care l-am putea introduce cu folos acum. Gândirea înseamnă acțiune „virtuală”, capacitatea de a testa lucruri în imaginație, capacitate care, din rațiuni biologice evidente, salvează vieți. Această capacitate nu este tipică oamenilor, dar în cazul nostru a ajuns la un nivel superior de dezvoltare. Să apelăm la un caz concret. Imaginează-ți situația următoare, derivată din experiența mea. Directorul mă ceartă în biroul lui. Mă simt tot mai furios. Răspunsul instinctiv este atacul afectiv. Acum *mă gândesc* la consecințele posibile. Așa că, în loc să mă arunc asupra lui, îmi reprim tendința instinctivă spre acțiune și îmi *imaginez* seria de alternative; îmi caut calea de ieșire în funcţie de ceea ce simt. În cele din urmă, mă opresc la o soluție satisfăcătoare: după ce ies din birou, când nu mă vede nimeni, îi sparg cauciucurile de la mașină. În felul acesta, îmi reduc FURIA fără să sufăr

consecințe cumplite. Vedem astfel, încă o dată, de ce stereotipurile comportamentale înnăscute trebuie să fie completate de învățarea din experiență, inclusiv de forma imaginară a experienței numită gândire. Când ne confruntăm cu frustrări reale, ce includ frecvent nevoi *conflictuale* (în acest caz, FURIE versus FRICĂ), soluțiile instinctive nu sunt suficiente. Dar de notat, din nou: dacă răspunsurile instinctive sunt completate de învățare nu înseamnă că ele sunt eliminate. Eu m-am hotărât să nu-l atac pe directorul meu, dar înclinația să o fac mi-a rămas și va apărea din nou în situații viitoare similare. (Acest exemplu nu e valabil numai în cazul oamenilor. Un câine poate că nu ar fi ajuns la o asemenea soluție, dar primatele fac tot felul de năzbâtii[13].)

Emoțiile precum FURIA nu sunt „simple" sentimente. Emoțiile joacă un rol fundamental în supraviețuire. Să ne imaginăm consecințele dacă nu am avea pretenţii la resursele disponibile și nu i-am împiedica pe alții să-și însușească partea noastră. Dacă nu am putea deveni frustrați, iritați sau furioși, nu am fi înclinați să luptăm pentru ce ne trebuie, iar în acest caz, mai devreme sau mai târziu, am muri. E ușor să trecem cu vederea funcția biologică a emoției în condițiile civilizate în care trăim astăzi. Dar trăim astfel (în așezări permanente cu legi artificiale ce reglementează comportamentul social) de doar aproximativ 12.000 de ani. Civilizația este o caracteristică recentă a existenței mamiferelor, nu a jucat niciun rol în modelarea creierilor noştri.

Gândirea conștientă necesită cortex, dar nu și sentimentele care o ghidează. Circuitul care mediază FURIA este aproape în întregime subcortical și, ca orice alte circuite afective, are destinația finală în PAG din trunchiul cerebral[14].

(4) FRICA. Dicotomia luptă/fugă arată că atacul afectiv nu este întotdeauna cea mai bună metodă de a-i face față unui adversar. Factorii contextuali ce separă lupta de fugă sunt înscriși în amigdală, care mediază și FURIA, și FRICA[15].

Majoritatea mamiferelor „știu” din prima zi că unele lucruri sunt inerent înfricoșătoare. Puii nou-născuţi ai rozătoarelor, de exemplu, încremenesc când sunt expuşi chiar la un singur fir de păr de pisică – deși nu au avut niciodată de-a face cu pisicile și nu știu nimic despre atitudinea lor față de șoareci. E ușor să înțelegem de ce stau lucrurile așa: dacă fiecare șoarece ar trebui să învețe din experiență cum să reacţioneze la pisici, atunci șoarecii ar dispărea ca specie. Din nou, vedem enorma valoare biologică a emoțiilor.

Noi, oamenii, ne temem de pericole precum înălțimile, locurile întunecate și creaturile care se târăsc spre noi și le evităm cu ajutorul acelorași instincte și reflexe ca ale altor mamifere: prin comportamente de încremenire și de fugă. Spre deosebire de reflexul vasovagal de leşin discutat mai sus, aceste comportamente de „evadare” sunt facilitate de respirația rapidă, de ritmul cardiac intensificat și de redirecționarea sângelui din intestine spre musculatura scheletului. (De unde și pierderea controlului sfincterelor asociată cu frica extremă.) Ca în cazul altor emoții, sentimentul conștient de frică îți spune dacă te îndrepți spre siguranță sau dacă te îndepărtezi de ea.

Un exemplu interesant este oferit de pacienta S.M., care avea aproape 30 de ani când a fost publicat cazul ei pentru prima dată. Suferea de sindromul Urbach-Wiethe, o afecțiune genetică rară ce duce la calcificarea bilaterală a amigdalei. Nu simțea teama. În consecință:

> [Ea] a fost victima a numeroase infracțiuni și incidente traumatice care i-au pus viața în pericol. A fost amenințată cu cuțitul și cu arma, a fost aproape omorâtă într-un incident de violență domestică și a primit amenințări explicite cu moartea în numeroase ocazii. În pofida faptului că multe dintre aceste situații îi puneau viața în pericol, S.M. nu manifesta niciun semn de disperare, urgență sau alte răspunsuri comportamentale ce ar fi asociate în mod normal cu asemenea incidente. Numărul disproporționat de mare de evenimente traumatice din viața lui S.M. este pus pe seama [...] incapacității sale severe de a detecta amenințările iminente din mediu și de a învăța să se țină departe de situațiile cu potențial de periculozitate[16].

Eu am examinat un mare număr de asemenea pacienți, întrucât sindromul Urbach-Wiethe are o incidență neobișnuit de mare într-un colţ izolat al Africii de Sud numit Namaqualand, care se întâmplă să fie în apropiere de locul unde m-am născut. (Gena defectă a fost adusă acolo de un colonist german, apoi s-a concentrat într-o comunitate izolată.) Mai ales visele lor mi s-au părut interesante – sunt scurte, simple și pline de aspirații. Una dintre pacientele pe care le-am studiat, al cărei soț era șomer, a visat: „Soțul meu și-a găsit un loc de muncă. Eram foarte fericită". Alta, mamă a unui copil cu dizabilități, a visat: „Fata mea putea să meargă. Eram foarte fericită". Un pacient al cărui tată murise a visat: „L-am văzut pe tata din nou. Eram așa de fericit!". Aceste vise sunt tipice pentru cei care suferă de sindromul Urbach-Wiethe, a căror imaginație neînfricată nu se așteaptă la niciun pericol în procesul

satisfacerii dorințelor lor[17].

Majoritatea oamenilor par că se nasc cu anumiți declanșatori specifici ai FRICII. Oare cum ar fi dacă fiecare dintre noi ar trebui să învețe din experiență ce se întâmplă când sari de pe o stâncă sau când pui mâna pe o viperă? Acesta este motivul pentru care ne tragem din strămoși care nu se simţeau înclinaţi să încerce. Cei care au încercat nu sunt strămoșii noștri pentru că ei nu au avut urmași. Avem toate motivele să fim recunoscători pentru această moștenire.

Dar trebuie să învățăm de ce *altceva* să ne temem. Învățăm din experiență – inclusiv cu ajutorul gândirii – că un rău asemănător ne poate fi cauzat și de alte lucruri în afară de căderea de la o înălțime ametitoare și de mușcăturile de șarpe. Evoluția nu ar fi avut cum să anticipeze prizele electrice și șocul pe care îl produc, de exemplu, dar ele sunt la fel de periculoase ca viperele. De asemenea, trebuie să învățăm ce altceva *să facem* când ne este frică, pentru a ne completa răspunsurile instinctive. Nu este adaptare să încremenești în fața a ceva care te sperie sau să fugi de acel ceva, tot așa cum nu este adaptare să ataci pe oricine te frustrează. Ar trebui să fie deja clar ce rol joacă sentimentele conștiente în procesul învățării – ne spun ce funcționează și ce nu înainte să fie prea târziu și, astfel, ne ajută să rămânem în viață.

Condiționarea prin frică relevă alte lucruri importante despre ce este conștient și ce nu. Una dintre trăsăturile ei speciale este „învățarea printr-o singură expunere” – nu trebuie decât să-ți bagi degetul într-o priză *o singură dată* ca să nu mai repeți acțiunea. E ușor să înțelegem de ce: ai avut noroc că ai supraviețuit prima dată, așa că de ce să repeți experiența? Totuși, la fel ca în cazul tuturor celorlalte mecanisme biologice

din spatele emoțiilor, nu e nevoie să cunoști acest lucru într-un sens „declarativ” – condiționarea se întâmplă pur și simplu automat, iar asta deoarece condiționarea prin FRICĂ nu necesită implicarea cortexului. Ea poate să aibă loc în prima copilărie, chiar înainte de maturizarea hipocampului (structura corticală responsabilă cu fixarea amintirilor declarative pe termen lung). Din acest motiv, ca în cazul lui Claparède, multor oameni sănătoși din punct de vedere neurologic le este frică de anumite lucruri fără să știe de ce.

Specialiștii în cogniție atribuie acest lucru învățării „inconștiente”, dar, după cum am spus deja, o fac doar pentru că neglijează afectul. E adevărat că mulți oameni nu conștientizează motivele pentru care se tem de diverse lucruri, dar sunt foarte conștienți de sentimentele asociate. Trebuie doar să simţi pentru a ghida comportamentul voluntar. Să ne amintim de experimentul cu succesiunea rapidă a unor cuvinte, descris mai sus: dacă termeni precum „criminal” și „violator” sunt asociaţi subliminal cu chipul A, iar „grijuliu” și „generos” cu chipul B, subiecții cercetării vor simți o preferință când li se va cere, mai târziu, să facă o alegere, deși nu pot spune de ce. Această alegere e ghidată de „instinct”, dar sentimentele rămân cu ușurință nerecunoscute, așa că sunt descrise din punct de vedere cognitiv prin cuvinte precum „a ghici”[18].

Astfel se explică o mare parte dintre nelămuririle din jurul „emoțiilor inconștiente” din știința cognitivă. Nu atât *emoțiile* sunt inconștiente, cât lucrurile *cognitive* la care se referă. După cum am văzut mai sus în legătură cu gândirea, cu siguranță ajută să știi la ce se referă sentimentele tale, dar această cunoaștere nu este esențială. De fapt, uneori este mai bine să *nu* gândești înainte să acționezi, nu în ultimul rând pentru că

gândirea presupune timp.

Aceleași afirmații sunt valabile şi în cazul condiționării prin frică. Odată ce ai învățat să te temi de ceva – mai ales dacă nu știi în mod conștient motivul –, asocierea este aproape ireversibilă. În memorabila formulare a lui Joseph LeDoux, amintirile legate de frică sunt „de neuitat”[19]. Asta ne spune lucruri importante despre memoria inconștientă în general, pe care le voi rezuma mai jos. Deocamdată, menționez doar că amintirile „nondeclarative” (precum cele emoționale și procedurale) sunt greu de uitat, din același motiv pentru care sunt inconștiente: conțin mai puțină incertitudine (adică sunt mai generalizabile) și, prin urmare, mai puțin susceptibile de revizuire contextuală. Acesta este modul în care comportamentele dobândite devin stereotipe și automate. În măsura în care obiectivul cunoașterii este să înveți cum să-ți satisfaci nevoile în lume, automatizarea este idealul învățării.

(5) PANICA/MÂHNIREA. Anxietatea de separare este diferită de FRICĂ. Apare iniţial în fragedă copilărie, când devii atașat instinctiv de mama ta (sau de principalul tău îngrijitor). Spre deosebire de condiționarea prin frică, dar din motive de natură biologică la fel de întemeiate, acest proces durează aproximativ șase luni: o singură situație de îngrijire afectuoasă nu este suficientă pentru a arăta că te poți baza pe o anumită persoană pentru totdeauna.

Când mamiferele sunt separate de ființele de care s-au atașat, se desfășoară o secvenţă stereotipă, începând cu un comportament de „protest”, urmat de „disperare”. Etapa protestului este caracterizată de senzații de panică, însoțită de vocalizări ale detresei și comportamente de căutare. Panica se

combină frecvent cu furia, care evocă alt conflict, de data asta între PANICĂ/MÂHNIRE și FURIE. Prima emoție te face să vrei să-ți ții îngrijitorul aproape pentru totdeauna, pe când cealaltă te face să vrei să-l distrugi. Vinovăția, o emoție *secundară* care inhibă FURIA, este de obicei rezultatul învățat. Acesta este un bun exemplu al modului în care emoțiile secundare (precum vinovăția, rușinea, invidia și gelozia) apar din situații conflictuale. Spre deosebire de emoțiile de bază, acestea sunt constructe învățate – hibrizi între emoție și cogniție (după cum arată cercetarea lui Barrett).

Etapa disperării este caracterizată de sentimente de deznădejde – literalmente de „renunțare". Explicația standard este: dacă plânsul și căutările puiului separat nu duc repede la reuniune, atunci costurile potențiale ale alertării prădătorilor cu privire la starea lui vulnerabilă încep să depășească beneficiile. De asemenea, dacă puiul se îndepărtează prea mult de adăpost, scad șansele să-l găsească mama când se întoarce. Astfel, din punct de vedere statistic, cum s-a întâmplat cu sincopa vasovagală, renunțarea (deși poate fi foarte dureroasă) devine strategia de supraviețuire moștenită.

Iată o descriere clasică a cascadei de separare la copii, făcută de psihanalistul John Bowlby:

> [Protestul…] poate începe imediat sau poate fi amânat; durează de la câteva ore la o săptămână sau mai mult. În acest timp, copilul mic pare grav tulburat de faptul că și-a pierdut mama și încearcă să o regăsească prin utilizarea deplină a resurselor sale limitate. Adesea plânge tare, își zgâlțâie pătuțul, se aruncă în toate părțile și se uită cu nerăbdare spre orice semn sau sunet care s-ar putea dovedi

că este mama lui absentă. Întregul lui comportament sugerează o expectanţă puternică că ea se va întoarce. Între timp, tinde să respingă orice alte persoane care se oferă să facă ceva pentru el, deși unii copii se vor agăța cu disperare de o asistentă medicală.

[Disperarea…] urmează protestului, preocuparea copilului față de mama absentă este încă evidentă, deși comportamentul lui sugerează tot mai multă deznădejde. Mișcările fizice active se diminuează sau încetează și e posibil să plângă monoton sau intermitent. Copilul e retras și inactiv, nu le cere nimic oamenilor din mediul său și pare să fie într-o stare de doliu profund[20].

Bineînțeles, cea din urmă stare este similară depresiei, care este însoțită frecvent de vinovăție. În consecință, Panksepp și alții (inclusiv eu) au aplicat această explicație a mecanismelor cerebrale ale PANICII/MÂHNIRII pentru a dezvolta tratamente noi ale tulburărilor de dispoziție[21]. Din punct de vedere chimic, tranziția de la „protest” la „disperare” este mediată de niște peptide numite opioide, care întrerup dopamina (vezi efectele lor în cazul descris mai jos, p. 132). Din acest motiv, depresia este caracterizată de contrariile în oglindă ale senzațiilor ce caracterizează CĂUTAREA[22]. Traiectoria anatomică a componentei PANICĂ a acestui sistem coboară din girusul cingular anterior în PAG, locul unde se termină toate circuitele emoțiilor[23]. (Voi explica mai jos cum de toate ciclurile afective se termină *și* încep în PAG din trunchiul cerebral.)

E interesant că acest circuit mediat de opioide a evoluat din sistemul analgezic mai vechi al creierului – suferința mintală a pierderii este o elaborare a mecanismelor corporale pentru

durerea senzorială[24]. Acesta este un bun exemplu al tranziției unitare existente în natură între afectele senzoriale ce salvează viața și trăirile emoționale. Emoțiile nu au nimic „fictiv”. Sentimentele dureroase asociate cu separarea și pierderea – cuplate cu învățarea din experiență – joacă un rol cauzal în asigurarea supraviețuirii mamiferelor și păsărilor, care *au nevoie* de îngrijitori. Acest fapt e valabil și după copilărie: circuitele cerebrale descrise mai sus mediază legăturile de atașament de-a lungul vieții, după cum mediază, din păcate, şi multe alte forme de dependență.

(6) GRIJA este cealaltă latură a atașamentului – nu numai că avem nevoie noi înșine de grijă afectuoasă, ci trebuie să avem grijă și de cei mici, mai ales de progeniturile noastre. Așa-numitul instinct matern există în noi toți, dar nu în același grad, deoarece este mediat de substanțe întâlnite la niveluri superioare (în general) la femei: estrogen, prolactină, progesteron și oxitocină – toate crescând semnificativ în timpul sarcinii și al nașterii[25]. Remarcabilă este și suprapunerea dintre chimia cerebrală și circuitele pentru GRIJĂ, PANICĂ/MÂHNIRE și DORINȚA SEXUALĂ tipică femeilor[26]. Chiar și numai aceste fapte pot să explice de ce depresia este mult mai întâlnită (de aproape trei ori mai frecvent) la femei decât la bărbați. Aproximativ 80% dintre persoanele de sex feminin știu cumva din copilărie că este „bine” să-i legene pe copii înspre partea stângă a liniei mediane a corpului, iar bărbații tind să descopere (instinctiv) acest lucru după ce devin tați[27]. Totodată, chiar și băieți complet neexperimentați știu ce să facă de obicei când plânge un bebeluș. Nu îl împung cu degetele și nici nu îl ridică de un picior ca să vadă dacă ajută la ceva, ci știu (prezic

în mod instinctiv) că este „bine” să-l țină aproape și să-l legene în timp ce scot sunete liniștitoare.

Și totuși, după cum învață fiecare părinte, aceste lucruri nu sunt suficiente. Ca să crești cu succes un copil până devine adult, e nevoie de mult mai mult decât instinct. Așadar, ca şi în cazul celorlalte emoții, trebuie să învățăm din experiență ce este de făcut în numeroasele situații neprevăzute care apar. Ca şi în cazul celorlalte emoții, deciziile noastre în acest sens sunt ghidate de sentimente (de grijă și preocupare) – care ne spun dacă lucrurile merg bine sau rău. Un alt motiv pentru care pulsiunea de a îngriji nu e suficientă este că nu simțim *numai* dragoste față de copiii noștri, după cum va confirma orice părinte. Conflictele ce rezultă de aici trebuie rezolvate prin procese cognitiv-emoționale hibride.

Învățarea modului de a împăca flexibil diversele nevoi emoționale stă la baza maturității și sănătății mintale. De exemplu, să ne gândim la relațiile de cuplu de durată, care necesită integrarea judicioasă a DORINȚEI SEXUALE cu atașamentul de tip PANICĂ/MÂHNIRE specific copiilor (să ne gândim la complexul virgină/prostituată, ce apare din incapacitatea de a împăca trăirile sexuale cu cele de afecțiune). Și legăturile afective sunt dificil de împăcat cu sistemul dinamic al CĂUTĂRII (să ne gândim la fiorul pe care ni-l dă noutatea), nu doar frustrările inevitabile ce cauzează FURIA (de unde și ubicuitatea conflictelor domestice), ceea ce intră în conflict cu preocupările GRIJII afectuoase ș.a.m.d. Menținerea relațiilor pe termen lung este doar un exemplu al numeroaselor provocări cu care se confruntă inima omului. Ca să gestionăm problemele vieţii, folosim emoțiile pe post de busolă. Ceea ce simţim ghidează orice învățare din experiență, în formele diferite pe

care le-am prezentat. Dar biologia ne oferă în ajutor încă o pulsiune:

(7) JOACA. *Avem nevoie* să ne jucăm. Joaca este mediul prin care se revendică și se apără teritorii, se formează ierarhii sociale și se stabilesc și se mențin granițele în interiorul și în afara grupului.

Oamenii sunt adesea surprinși să afle că e o pulsiune biologică, dar toţi puii de mamifere se implică în joacă dinamică și destul de dură. Dacă nu își iau porția de joacă într-o zi, vor încerca să și-o recupereze a doua zi. Știm ce înseamnă încăierarea în joacă, deși forma pe care o ia variază puţin de la o specie de mamifere la alta. O sesiune de joacă începe cu un anumit gest sau o anumită postură în semn de „invitație”, apoi, dacă este acceptată, începe joaca. Animalul sau copilul îl fugărește pe celălalt plin de exuberanță, apoi se opresc ca să se ia la trântă sau să se gâdile, fiecare stând deasupra, pe rând – toate acestea însoțite de hohote de râs sau de vocalizarea echivalentă, în funcție de specie (chiar și șobolanii „râd”)[28]. Apoi sar din nou în picioare, fugărindu-se în direcția opusă. Trăirea emoţională asociată este la fel de universală: se numește distracție.

Copiii iubesc joaca. Totuși, empiric, majoritatea episoadelor de joacă se încheie cu lacrimi. Acest lucru oferă un indiciu important pentru a înţelege despre ce e vorba de fapt, dintr-o perspectivă biologică: e vorba de *găsirea limitelor* a ceea ce este tolerabil și permisibil din punct de vedere social. Când jocul nu mai este distractiv pentru unul dintre participanți, adesea din cauză că hotărăște că nu ești „corect”, acesta nu mai vrea să se joace. Și-a atins limita. Marcarea unor asemenea limite este

esențială pentru formarea și menținerea unor grupuri sociale stabile. Iar supraviețuirea unui grup este importantă pentru supraviețuirea fiecărui membru al grupului la speciile sociale cum este a noastră.

În acest sens, un criteriu important este *dominarea*. În orice situație de joacă, unul dintre participanți își asumă rolul de conducere, iar celălalt se supune. Așa e distractiv pentru amândoi, atât timp cât cel dominant nu insistă să fie la comandă *tot* timpul. Raportul acceptabil al schimbării rolurilor pare să fie în jur de 60:40. Conform „regulii 60:40” a *reciprocității*, jucătorul supus continuă să joace atât timp cât i se oferă suficiente oportunități să preia conducerea.

Acest lucru relevă o a doua funcție a JOCULUI: stabilirea unor ierarhii sociale. În consecință, încăierarea în joacă lasă locul (odată cu dezvoltarea) unor jocuri mai organizate și deschis competitive. Desigur, joaca nu se limitează la simularea unor încăierări. Noi, oamenii, ne implicăm în diverse jocuri de rol, în care participanții încearcă diferite roluri sociale (de exemplu, de-a mama și copilul, profesorul și elevul, doctorul și pacientul, hoțul și vardistul, cowboyul și indianul[29], regele și tâlharul – de observat omniprezentele ierarhii de status şi de putere). Nu știm ce se petrece în imaginația altor mamifere când se joacă, dar putem presupune cu siguranță că și ele „încearcă” diferite roluri sociale, învățând astfel ce este acceptabil și ce este inacceptabil.

Acest lucru sugerează o a treia funcție biologică pentru JOACĂ. Îți cere (și te condiționează) să ții cont de sentimentele altora. Dacă nu ții cont, ceilalți vor refuza să se joace cu tine, iar tu vei fi lipsit de plăcerea enormă pe care ți-o aduce joaca. Deşi cel care face bullying acaparează toate jucăriile, el va fi privat

de toată distracția. Se pare că acesta e motivul pentru care JOACA a evoluat (și pentru care este însoţită de atâta plăcere): promovează formațiuni sociale viabile. Pe scurt, este un vehicul important pentru dezvoltarea *empatiei*[30].

Episoadele de joacă încetează abrupt atunci când își pierd calitatea de „ca și cum”. Dacă îți încui surioara undeva și arunci cheia, atunci nu numai că ai încălcat regula 60:40, ci nici nu mai e vorba de *jocul* de-a hoții și vardiștii, ci de faptul că îți ții sora captivă. Cu alte cuvinte, acum comportamentul vostru reciproc este guvernat de FRICĂ sau de FURIE, și nu de JOACĂ. Aceeași afirmație e valabilă și în cazul celorlalte jocuri enumerate mai sus. „De-a doctorul și pacientul” este un joc până ce se transformă în act sexual real, apoi e guvernat de DORINȚA SEXUALĂ. Faptul că JOACA rămâne, cum s-ar spune, între toate celelalte emoții instinctuale – experimentându-le și învățându-le limitele – este probabil motivul pentru care nu se poate identifica un singur circuit cerebral pentru JOACĂ. Probabil că sunt implicate toate[31]. Totuși, cine se îndoiește că joaca este un instinct primar ar trebui să citească minunata carte scrisă de Sergio și Vivien Pellis *The Playful Brain*[32].

Nu ne place întotdeauna să recunoaștem că oamenii, asemenea altor mamifere, își revendică teritorii și formează ierarhii sociale cu reguli clare în mod natural. (Regulile ce guvernează comportamentul primatelor sunt remarcabil de complexe.) Structura familiilor, a clanurilor, a armatelor și chiar a țărilor – aproape a oricărui grup social – este fără îndoială ierarhică și teritorială, iar asta se vede în orice perioadă istorică. Pe cât statutul social al unui individ în interiorul grupului este mai înalt, pe atât mai mult acces are individul la resursele din teritoriul controlat de grup. Această

observație nu ține de preferința personală, ci e un fapt. Dacă nu recunoaștem asemenea fapte, nu ne putem ocupa de ele. Faptul că există pulsiuni emoționale nu înseamnă că nu le putem controla – că suntem obligați să respectăm „legea junglei” –, dar ne asumăm un risc dacă le ignorăm.

E ușor de observat modul în care JOACA, mai ales, creează reguli sociale. Ele reglează comportamentul de grup, protejându-ne astfel de excesele nevoilor noastre individuale. E ușor de observat și modul în care regulile sociale încurajează forme complexe de comunicare și, prin urmare, felul cum contribuie la apariția gândirii simbolice. Calitatea de „ca și cum” a circuitului JOACĂ sugerează că joaca ar putea fi chiar precursorul biologic al gândirii în general (adică a acțiunii virtuale versus acțiunea reală; vezi mai sus). Unii cercetători consideră că și *visul* este JOACĂ nocturnă, adică testarea emoțiilor instinctuale într-o lume „ca și cum”. E interesant că, în tulburările comportamentale în somnul paradoxal, când paralizia motoare ce însoțește în mod normal visele dispare din cauza unor leziuni la nivelul trunchiului cerebral, pacienții (și animalele din cadrul unor experimente) *joacă* fizic diferitele stereotipuri instinctuale – de exemplu, fugă, încremenire, atac de prădător și atac afectiv.

Sper că acum se înțelege că, deși afectele corporale au o anumită intensitate și o imediatețe ce le fac să pară într-un anumit sens de neratat, afectele emoționale acționează și ele prin sentimente conștiente. Deși nu le recunoaștem întotdeauna ca atare, ele reglează mai tot comportamentul nostru voluntar prin diversele lor imbolduri şi izbucniri interioare. Comportamentul voluntar constă, în esență, în a face alegeri de

tip aici-și-acum. Cum să faci alegeri fără ca ele să se bazeze pe vreun sistem de evaluare ce îți spune care opțiune este mai bună sau mai proastă? Cu aceste valori contribuie la comportament faptul de a simţi.

Totuși, deoarece stările noastre emoționale din fundal nu sunt recunoscute întotdeauna de creierul *cognitiv* și, astfel, nu sunt întotdeauna declarate într-un mod autoreflexiv, noi nu vedem tiparul general până când nu privim în urmă și punem lucrurile cap la cap. Aici, când spun „noi”, nu mă refer numai la nespecialiști, care au tot dreptul să nu cunoască datele experimentale. Deși e adevărat că trecem cu vederea prea ușor rolul esențial al sentimentelor în viața de zi cu zi, de fapt mă refer la opiniile general acceptate în științele cognitive contemporane. Specialiștii în științe cognitive de obicei trec cu vederea sentimentele.

Dar, după cum se va vedea, dintr-o perspectivă neurologică, în capitolul următor, orice descriere științifică a conștiinței ce ignoră rolul fundamental al sentimentelor ratează evenimentul principal.

*1 Panksepp (1998). Nu folosesc termenul „homeostatic” în accepțiunea lui Panksepp pentru subtipul interoceptiv de afect corporal deoarece, după cum vom vedea imediat, toate afectele sunt homeostatice. Din acest motiv, utilizarea mai restrânsă a termenului la Panksepp poate fi derutantă.

6

Sursa

O ipoteză fundamentală în neuropsihologie – bazată pe metoda clinico-anatomică – este că, dacă o anumită funcție psihică e îndeplinită de o anumită regiune cerebrală, atunci o leziune totală la nivelul ariei respective trebuie să ducă la pierderea totală a acelei funcții. După cum am arătat, când e vorba de conștiință, cortexul nu trece acest test. Dar situația este și mai proastă pentru teoria conform căreia conștiința apare exclusiv în cortex: unele leziuni *din altă parte* a creierului o anulează complet – și încă leziuni foarte mici.

Fiziologii Giuseppe Moruzzi și Horace Magoun au stabilit în urmă cu peste 70 de ani că conștiința se pierde la pisici în urma unor mici incizii care deconectează cortexul de nucleul „reticulat” (ca o plasă) al trunchiului cerebral[1]. Acest nucleu are probabil vreo 525 de milioane de ani, întrucât este comun tuturor vertebratelor – de la pești la oameni. De la descoperirea lui de către Moruzzi și Magoun, cercetătorii confirmă la tot felul de specii că leziuni relativ mici la nivelul acestui nucleu – cunoscut sub denumirea tehnică de sistem reticulat activator – cauzează comă. De exemplu, David Fischer și colegii lui au identificat recent la pacienții umani cu accident vascular în

trunchiul cerebral o regiune mică „specific comatoasă” de doi milimetri pătrați, în partea superioară a tegmentului mezopontin (vezi figura 1)[2].

Există două explicații posibile. Prima este că acest centru al trunchiului cerebral cu o mare densitate de noduli e locul unde apare conștiința: izvorul ascuns al minții, sursa esenței sale. Şi sunt de această părere, ca și Jaak Panksepp. A doua explicație ar fi că acesta e precum cablul de alimentare pentru un televizor: necesar, dar nu suficient și nu prea te ajută dacă vrei să înțelegi cum funcționează televiziunea. Aceasta este opinia predominantă.

Să presupunem că a doua opțiune este adevărată. Prin stimularea trunchiului cerebral, ne-am putea aștepta să putem să pornim și să oprim conștiința. În cel mai bun caz, am putea să o atenuăm în anumite feluri, așa cum o fluctuație de tensiune face ecranul televizorului să pâlpâie. Nu ne-am aștepta să refacă din mers emisia curentă. Și totuși, un electrod implantat într-un nucleu reticulat din trunchiul cerebral al unei femei de 65 de ani (pentru tratarea bolii Parkinson) a generat următorul răspuns remarcabil:

> Chipul pacientei a exprimat o tristețe profundă în mai puţin de cinci secunde [...]. Deși era încă alertă, pacienta s-a lăsat spre dreapta, a început să plângă și a comunicat verbal sentimente de tristețe, vinovăție, inutilitate și deznădejde, precum „Mă prăbușesc în capul meu, nu mai vreau să trăiesc, nu vreau să mai văd, să mai aud, să mai simt nimic...”. Când a fost întrebată de ce plânge și dacă o doare ceva, a răspuns: „Nu, m-am săturat de viață, m-am săturat... Nu mai vreau să trăiesc, mi-e silă de viață... Totul e inutil,

> mereu mă simt inutilă, lumea asta mă sperie”. Când a fost întrebată de ce e tristă, a răspuns: „Am obosit. Vreau să mă ascund într-un colț… Îmi plâng de milă, desigur… N-am nicio speranță, de ce vă mai deranjez?” […]. Depresia a dispărut în mai puțin de 90 de secunde după oprirea stimulării. În următoarele cinci minute, pacienta s-a aflat într-o stare ușor hipomaniacală, a râs și a glumit cu examinatorul, trăgându-l în glumă de cravată. Și-a amintit întregul episod. Stimularea [într-o altă regiune cerebrală, care era adevărata țintă a electrodului] nu a generat același răspuns psihiatric[3].

Această pacientă nu avea istoric de simptome psihiatrice de vreun fel.

Aceste afirmații sunt valabile și în cazul stimulării sau blocării *chimice* a acestor nuclei din centrul trunchiului cerebral. Majoritatea antidepresivelor – activatori ai serotoninei – acționează asupra unor neuroni ai căror corpi celulari se află într-o regiune a sistemului reticulat activator numită nucleii rafeului (vezi figura 1). Acolo e „sursa” serotoninei. Antipsihoticele – blocanți ai dopaminei – acționează asupra unor neuroni a căror sursă e în altă parte a sistemului reticulat activator: aria tegmentală ventrală (vezi figura 2). Afirmația e valabilă și în cazul anxioliticelor – multe dintre acestea blocând o substanță numită noradrenalină, care este produsă de neuroni ce își au sursa în *locus coeruleus complex* (tot figura 1), o altă parte a sistemului reticulat activator. Toți acești neuroni sunt aglutinați în centrul reticulat al trunchiului cerebral. Psihiatrii nu ar testa această regiune cerebrală dacă ea doar ar porni și ar opri conștiința. Dacă ea ar face doar acest lucru, atunci i-ar interesa numai pe anesteziști.

Așadar, a doua opinie trebuie să fie greșită.

Neuroimagistica funcțională a creierului în unele stări emoționale indică aceeași concluzie. Tomografia cu emisie de pozitroni în timpul stărilor de MÂHNIRE, CĂUTARE, FURIE și FRICĂ, de exemplu, arată că activitatea metabolică cea mai intensă se desfășoară în centrul trunchiului cerebral (și în alte zone subcorticale; vezi figura 9), iar cortexul arată *deactivare*. Imagistica prin rezonanță magnetică funcțională în timpul orgasmului relevă același lucru: activitatea hemodinamică ce se corelează cu această stare intens afectivă se localizează aproape exclusiv în mezencefal[4].

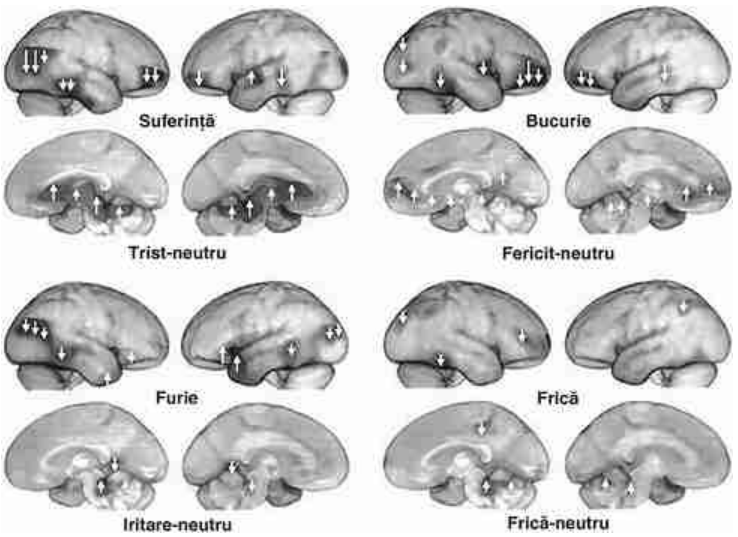


Figura 9. Imagini din tomografii cu emisie de pozitroni a patru stări emoționale (prin amabilitatea lui Antonio Damasio). Săgețile ascendente indică regiuni cu activitate *crescută*, iar săgețile descendente indică regiuni cu activitate *scăzută*. Zona evidențiată din imaginea „Bucurie” pare că arată activarea sistemului CĂUTĂRII.

Studiile asupra unor leziuni, stimularea cerebrală profundă, manipularea farmacologică și neuroimagistica funcțională indică aceeași concluzie: centrul reticulat al trunchiului cerebral generează *afecte*. Prin urmare, aparent, singura parte a creierului despre care știm că este necesară pentru activarea conștiinței în ansamblul său are o influență la fel de puternică asupra altei funcții psihice, şi anume sentimentele. În capitolul anterior, am arătat că sentimentele răzbat în întreaga experiență conștientă: indiferent ce altceva e menită să facă, una dintre principalele sarcini ale conștiinței pare să constea în a avea și a gestiona sentimente (care vin din *interiorul* nostru și ne reglează nevoile biologice). Dar acum pare că sursele neurologice ale afectului și ale conștiinței cel puțin se întrepătrund în mare măsură și, de fapt, s-ar putea să coincidă. Contrar opiniei empiriste clasice, conform căreia conștiința ne parvine prin simțuri, și contrar afirmației citate din Meynert ce se bazează pe această opinie, se pare că creierul *chiar* „radiază propria căldură”.

Cum ar trebui să numim mediul de bază, această misterioasă materie a minții ce pare să izvorască în interiorul nostru? Nu o putem numi „stare de veghe”, așa cum o numea Zeman, întrucât asta ne-ar impune să descriem visul drept un tip de trezie, ceea ce este absurd. Nu o putem numi nici „nivel” cantitativ, ca Moruzzi și Magoun, întrucât faptele descrise mai sus arată că ea are trăsături intens calitative.

Așadar, să încercăm al treilea termen folosit în literatura de specialitate: „activare”. Mi se pare un cuvânt bun, neutru. Atât starea de veghe, cât și visul implică activare; și nu exclude calitatea – ca în cazul termenului „nivel”. De fapt, „activare” sugerează categoric faptul de a simţi.

Dar ce *este* activarea? Am vorbit despre ea în termeni făţiş behavioriști – de exemplu, deosebirea dintre comă, starea vegetativă și starea de veghe cu reacții complete. De obicei, e măsurată cu ajutorul scalei de comă Glasgow: testarea răspunsului ocular al pacientului, a răspunsului verbal și a răspunsului motor la instrucțiuni și la durere. Totuși, ea poate fi definită și din punct de vedere psihologic.

O electroencefalogramă (EEG) produce reprezentări grafice ale activității electrice corticale. Când e lăsat în voia lui (adică este deconectat de la sistemul reticulat activator, chiar şi atunci când procesează inputuri senzoriale), cortexul produce tiparul undelor *delta*, o serie de unde cu amplitudine mare ce apar de aproximativ două ori pe secundă (adică o frecvență de 2 Hz). Când este stimulat de sistemul reticulat activator în absența unui input senzorial, cortexul produce de regulă ritmul *teta* (4-7 Hz)[5] sau ritmul *alfa* (unde desincronizate la frecvențe de 8-13 Hz; „desincronizat” înseamnă neregulat). Când prelucrează în mod activ informații din exterior, cortexul are de obicei ritm *beta* (unde desincronizate, de amplitudine foarte mică, cu frecvențe de 14-24 Hz) sau *gama* (unde de mică amplitudine, cu frecvențe foarte mari, de 25-100 Hz). Gama este ritmul asociat de regulă cu conștiința.

Astăzi, activarea fiziologică poate fi măsurată și cu ajutorul neuroimagisticii funcționale, ce ilustrează la propriu activitatea cerebrală prin topografierea unor ritmuri regionale de modificare a ratelor metabolice. Figura 3 ilustrează această tehnică, cu referire la diferite faze ale somnului. Rândul de jos din figură arată activarea somnului REM, care este de obicei (deși, după cum am arătat, nu exclusiv) asociat cu conștiința în

timpul visului și care își are sursa în trunchiul cerebral superior. Neuroimagistica unora dintre stările emoționale de bază ale conștiinței descrise mai sus (vezi figura 9) și ale orgasmului arată același lucru: activarea trunchiului cerebral.

Cortexul devine conștient numai în măsura în care este activat de trunchiul cerebral. Relația dintre cele două este ierarhică; conștiința corticală *depinde* de activarea din trunchiul cerebral. Acesta este motivul pentru care ritmul undelor delta prezentat primul de sus în figura 10 – care nu este asociat cu comportamentul conștient – este generat de activitatea corticală intrinsecă și tot din acest motiv ritmul gama prezentat ultimul în figura 10 – care este puternic asociat cu conștiința – poate fi impulsionat exclusiv de sistemul reticulat activator[6]. Și tot din acest motiv scăderea activării fiziologice ilustrată pe rândul de sus din figura 3 coincide cu atenuarea conștiinței și instalarea somnului, iar activarea intensificată ilustrată pe rândul de jos coincide cu reapariția conștiinței în timpul visului. Acestea nu sunt fapte controversate.

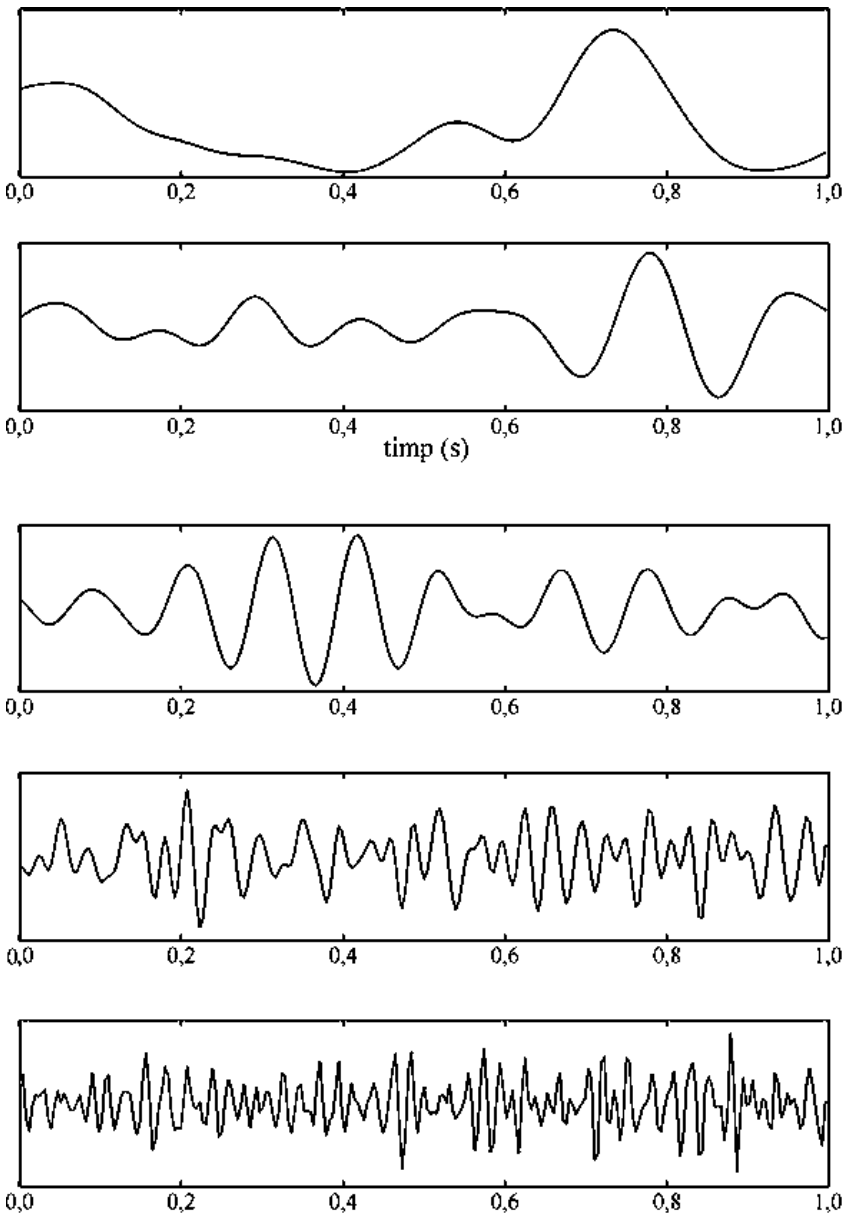


Figura 10. EEG-uri cu ritmuri tipice ale activității corticale. Ritmul cel mai puțin activat (delta) este primul de sus, iar ritmul cel mai activat (gama) ultimul. Ritmurile intermediare, de sus în jos, sunt teta, alfa și beta.

Să examinăm mai atent mecanismele cerebrale propriu-zise implicate. Pentru a face acest lucru, trebuie să vorbesc despre o deosebire fundamentală între două moduri în care neuronii

comunică între ei. Se dovedește că această deosebire este importantă pentru conștiință.

Majoritatea celor oarecum interesați de creier știu că neuronii transmit mesaje de-a lungul unor rețele complexe. Acest proces se numește *transmisie sinaptică*, deoarece implică transmiterea unor mesaje prin intermediul sinapselor, structurile prin care neuronii își transmit semnale (cuvântul „sinapsă” vine din termenul grecesc pentru „a prinde împreună”). Transmisia sinaptică folosește molecule numite *neurotransmițători*, care sunt trecute de la un neuron la următorul fie excitând neuronul postsinaptic, fie inhibându-l, în funcție de molecula în cauză (glutamatul și aspartatul sunt neurotransmițători excitatori, iar acidul gamaaminobutiric [GABA] este inhibitor). Dacă neuronul din aval este excitat de un val de neurotransmițători, își transmite propriile molecule următorilor neuroni din rețea. În caz contrar, nu transmite. Apoi, moleculele transmise se degradează rapid sau sunt luate înapoi în neuronul presinaptic, pentru a se limita durata efectului lor – un proces numit „recaptare”.

Transmisia sinaptică este țintită, binară (da/nu) și rapidă. Acest aspect al funcționării creierului amintește cel mai mult de calculul digital, astfel explicându-se poate de ce suscită atât de mult interesul specialiștilor în neuroștiințe care au înclinații spre calculatoare. Ea apare în tot sistemul nervos, inclusiv în cortex. Dar nu este intrinsec conștientă. Cu alte cuvinte, acest tip de neurotransmisie apare în cortex, indiferent dacă este conștientă sau nu. Și nu are aproape nimic de-a face cu activarea.

Puțini știu însă că transmisia sinaptică are loc sub influența

constantă a unui proces fiziologic complet diferit. Acest alt tip de activitate neuronală se numește *modulare postsinaptică*. Spre deosebire de transmisia sinaptică, modularea postsinaptică este dezordonată, chimică și foarte diferită de ce se întâmplă într-un computer obișnuit. Ea apare în mod endogen din sistemul reticulat activator (și din alte structuri subcorticale și chiar din unele structuri nonneurologice ale corpului) și are legătură cu activarea.

Jucătorii principali în acest proces sunt o clasă de molecule numite *neuromodulatori*. Spre deosebire de neurotransmițători, aceste molecule sunt răspândite în tot creierul – adică sunt eliberate în apropierea unor populații întregi de neuroni, și nu spre sinapse individuale[*1]. În loc să transmită mesaje de-a lungul unor „canale” specifice, ele inundă legături din rețea, reglând astfel „starea” generală a cortexului. De exemplu, cortexul se află în stări diferite pe rândurile de sus și de jos ale imaginilor din figura 3 (somn cu unde lente versus somn REM) și în cele patru stări emoționale ilustrate în figura 9 (MÂHNIERE versus CĂUTARE versus FURIE versus FRICĂ). În fiecare dintre aceste stări, el prelucrează diferit informațiile. Astfel, dacă cineva te strigă pe nume când dormi, reacționezi foarte diferit față de atunci când ești complet treaz[7]. De asemenea, gândește-te ce reacție ai față de un străin care se apropie dacă ești într-o stare de CĂUTARE versus FRICĂ: în prima stare poate că l-ai saluta pe străin și chiar ai începe o conversație cu el, iar în a doua stare probabil ai privi în altă parte, sperând că nu te observă.

Această deosebire dintre „canal” și „stare” este o prescurtare utilă pentru cele două moduri în care neuronii comunică între ei[8]. *Starea* cortexului afectează intensitățile diferențiale ale

transmiterii mesajelor ce continuă în interiorul *canalelor* sale; într-un anumit sens, starea reglează cât de „sonor” comunică diferitele canale (vezi figura 11). Acesta este motivul pentru care același sunet (de exemplu, cineva care te strigă pe nume) este generalizat în cortex în timpul stării de veghe, dar este izolat în cortexul auditiv în timpul somnului, iar prezența unui străin activează o rețea cerebrală în timpul stării de FRICĂ și alta în timpul CĂUTĂRII.

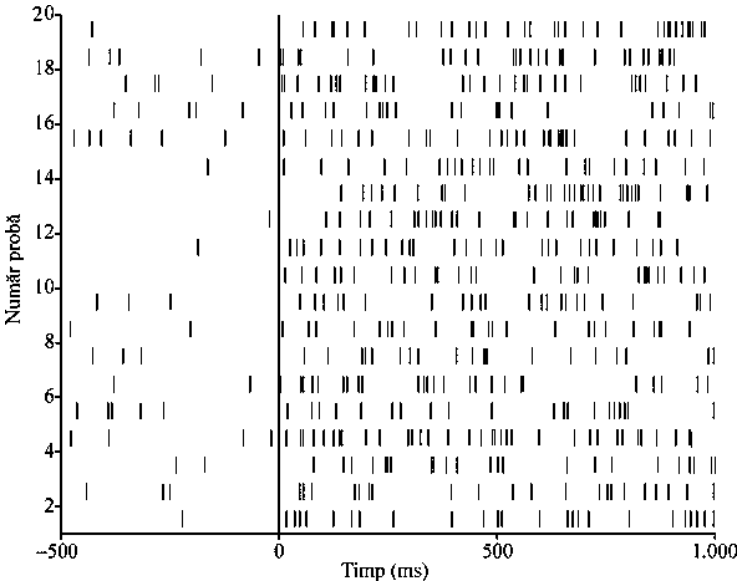


Figura 11. Imaginea de mai sus este o reprezentare grafică a măsurătorilor potențialelor de acțiune (*spike trains*) a 20 de neuroni (reprezentați pe axa verticală *y*) într-un interval de 1,5 secunde (reprezentat pe axa orizontală *x*) în timpul expunerii la un stimul vizual. Expunerea la stimul intervine la momentul 0 pe axa *x* (reprezentată de a doua linie verticală). În acest punct, neuronii, care au o frecvenţă de descărcare inițială de 6 Hz (în medie, când nu există stimul), își intensifică frecvenţa de descărcare medie la 30 Hz. Un *spike train* este o succesiune temporală în care un neuron se descarcă (= acțiune) și nu se descarcă (= latență). Acest lucru poate fi reprezentat ca o

secvenţă digitală de informații: „1” pentru o acțiune și „0” pentru o latență. De exemplu, un *spike train* ar putea fi reprezentat în cod ca „001111101101”. Aici, primele două zerouri reprezintă intervalul de latență dintre expunerea la stimul și prima acțiune. E important de menționat că frecvenţele de descărcare nu sunt determinate numai de stimul, ci ele apar dintr-o interacțiune între stimul, potențarea sau inhibarea pe termen lung a neuronilor (adică dacă stimulul este familiar sau nefamiliar) și nivelul curent al modulării lor. Ajustarea modulării postsinaptice este efectul activării asupra neuronilor. Din acest motiv, stimulul ilustrat aici ar putea să nu evoce niciun răspuns din partea acelorași neuroni atunci când activarea este modulată descendent.

Aceasta este esența a ceea ce numim activare. De observat însă că activarea corticală poate fi modulată atât ascendent, *cât și descendent*, până la suprimarea totală a transmisiei, după cum se întâmplă în fiecare noapte când ne culcăm (motiv pentru care unii fiziologi preferă termenul „modulare” în loc de „activare”). Prin urmare, activarea determină ce impulsuri sinaptice vor fi transmise și cât de puternic, ca în exemplul în care ești strigat pe nume când dormi versus când ești treaz.

Transmisia sinaptică este binară (pornit/oprit, da/nu, 1/0), dar neuromodularea postsinaptică *ierarhizează* probabilitatea ca un anumit set de neuroni să se descarce. Modifică şansele statistice ca ceva să se întâmple în interiorul lor. Această ajustare probabilistă și analoagă a frecvenţelor de descărcare se face prin receptori aflați în mai multe locuri de-a lungul unui neuron. Spre deosebire de neurotransmițători, neuromodulatorii au efecte relativ lente și de durată – nu

numai pentru că substanțele se mențin mai mult timp, ci și pentru că, în cazul canalelor care se descarcă mai frecvent, crește probabilitatea de a se descărca în viitor. Dacă activezi o anumită parte a rețelei, ea va rămâne oarecum activată până când este modulată descendent. Acest lucru influențează neuroplasticitatea și constituie o parte importantă a modului în care funcționează învățarea. Stările de activare înregistrează diversele lecții pe care le învățăm mai profund în canalele creierului. Astfel, de exemplu, probabilitatea să-ți amintești de o călătorie este mai mare dacă încerci cu nerăbdare să găsești o destinație nefamiliară decât dacă tot călătorești în același loc din obișnuință, pe pilot automat.

De unde provin neuromodulatorii? Din tot corpul, inclusiv din glandele hipofiză, suprarenală, tiroidă și sexuale (care produc diverși hormoni) și din hipotalamus (care produce numeroase peptide). Dar sursa principală a „activării" din punctul de vedere al creierului este sistemul reticulat activator. Activarea reticulată a trunchiului cerebral eliberează cei mai cunoscuți cinci neuromodulatori: *dopamina* (având sursa principală în aria tegmentală ventrală și în substanța neagră), *noradrenalina* (având sursa principală în *locus coeruleus complex*), *acetilcolina* (având sursa principală în nucleii tegmentului mezopontin și ai prozencefalului bazal), *serotonina* (având sursa principală în nucleii rafeului) și *histamina* (având sursa principală în corpii mamilari ai hipotalamusului). Nu întâmplător unele dintre aceste substanțe au reprezentat subiecte importante de discuţie în capitolele anterioare. Fiecare dintre molecule și receptorii lor aferenți, care au multe subtipuri, este responsabil pentru un aspect diferit al activării modulatoare. Pe lângă acestea cinci, există numeroase altele –

în principal hormoni și peptide cu acțiune lentă (în creier sunt peste o sută) – care modulează sisteme neuronale cu grad mare de specificitate.

Efectul tuturor acestor modulatori este determinat de prezența sau absența receptorilor relevanți și a subtipurilor de receptori. Cu alte cuvinte, deși neuromodulatorii sunt răspândiți în întregul creier, ei influențează numai celulele care au receptorii relevanți. Nu trebuie să ne gândim la activare ca la un proces brut sau nediscriminatoriu. Ea are foarte multe aspecte și se desfășoară sub multe dimensiuni, atât spațial, cât și temporal. Mesajele potențiale purtate de neuromodulatori circulă prin numeroasele rețele cerebrale, dar (asemenea analgezicelor) sunt utilizate numai acolo unde este nevoie de ele. De exemplu: în timp ce citești acest text, tot felul de stimuli de fond sunt înregistrați de sistemele tale senzoriale fără să le acorzi atenție, dar, dacă ești mamă și nou-născutul tău începe să plângă, acest lucru devine prioritar față de atenția pe care o acorzi cărții și imediat ți-o redirecționezi către bebeluș. Asta se datorează nivelurilor crescute ale peptidelor și hormonilor specifici – estrogen, progesteron, prolactină și oxitocină – care îți circulă prin creier când ai un bebeluș, schimbându-ți starea (adică activează sistemul GRIJĂ al creierului).

Neuromodulatorii pot să moduleze (ascendent sau descendent) numai semnale care sunt prezente – cu alte cuvinte, canale active în momentul respectiv. Ei sunt eliberați în mod difuz, dar influențează numai neuronii care (1) au receptorii relevanți și (2) sunt activi în momentul respectiv. Același modulator are efecte diferite asupra unor clase diferite de receptori, în locuri diverse. Se obține mai multă specificitate de către modulatorii eliberați nu numai de trunchiul cerebral

(și de sistemul endocrin), ci și în circuitele neuronale. Cei mai cunoscuți cinci modulatori ai activării pe care i-am enumerat mai sus își au sursa principală în trunchiul cerebral reticulat. Hormonii își au sursa în alte părți ale corpului și ajung la creier în principal prin fluxul sangvin. Modulatorii mai specifici sunt peptidele, multe dintre ele avându-și sursa în hipotalamus. Fiecare peptidă face ceva diferit, din nou în funcție de tipul și localizarea receptorului. Multe dintre ele, care modulează emoțiile de bază în moduri foarte precise, sunt enumerate în notele de final ale capitolului 5.

Până acum în acest capitol m-am ocupat de două întrebări: *de unde* vine activarea din punct de vedere anatomic și *cum* apare ea din punct de vedere fiziologic. Răspunsurile sunt că activarea este generată în principal, dar nu exclusiv în trunchiul cerebral și în hipotalamus și activează prozencefalul modulând neurotransmisia. Pentru a reaminti din capitolul 5, *motivul pentru care* apare activarea este că ea răspunde unor solicitări endogene de travaliu asupra psihicului. Aceste solicitări iau forma multor semnale „de eroare” ce converg în centrul creierului. În majoritatea cazurilor, asemenea semnale sunt gestionate în mod automat și inconștient, solicitându-se răspunsuri conștiente numai în situații în care răspunsurile automate nu sunt suficiente. Iar astfel ajung la întrebarea principală a capitolului de faţă: unde apare această trecere aparent magică de la reflexul automat la sentimentul volitiv?

Ceea ce urmează să descriu și să explic nu este definitiv, din cauză că faptele încă nu sunt cu totul clare. Cunoașterea actuală din neuroștiințe este însă suficient de avansată pentru a face o descriere în linii mari a imaginii de ansamblu. Pionierul acestor

cercetări a fost Jaak Panksepp, urmat de Antonio Damasio. Damasio are darul de a vedea imaginea de ansablu, dar specialistul în neuroștiințe căruia consider că îi suntem cei mai îndatorați aici, cu privire la modul în care creierul face saltul misterios de la reflexul automat la acțiunea voluntară, este Björn Merker. El studiază activarea trunchiului cerebral și mecanismele tropismului la o mare diversitate de specii de vertebrate, inclusiv păsări, rozătoare, feline și primate.

Aparent, trecerea de la starea de veghe vegetativă la activarea afectivă depinde de integritatea unui conglomerat mic, foarte compact de neuroni din jurul canalului central al mezencefalului – *substanța cenușie periapeductală* (PAG), unde converg *toate* circuitele cerebrale afective (vezi figura 6 pentru localizarea acestei structuri). Prin urmare, pacienții cu leziuni localizate în PAG „au privirea pierdută într-un spațiu psihoafectiv gol”[9]:

> Leziunile extinse la nivelul PAG [cauzează] o deteriorare spectaculoasă a tuturor activităților conștiente [...]. De exemplu, primele studii în care au fost inserați electrozi pentru producerea unor leziuni din al patrulea ventricul ascendent prin apeduct spre limita caudală a diencefalului au evidențiat deficiențe impresionante ale conștiinței la pisici și maimuțe, după cum au arătat *eșecul lor de a manifesta orice comportament intențional aparent* și *lipsa generală de răspuns la stimuli emoționali*[10]. Deși unele forme de leziuni în oricare alte zone cerebrale superioare pot să afecteze „instrumentele [cognitive] ale conștiinței”, de regulă ele nu deteriorează baza intenționalității în sine. Leziunile PAG reușesc acest lucru cu cea mai mică

distrugere absolută a țesutului cerebral[11].

Starea descrisă este „starea vegetativă”. Principala diferență dintre aceasta și comă este menținerea stării de veghe. Dar ritmul circadian somn/veghe este doar o altă funcție autonomă[12]. Din acest motiv, starea vegetativă se numește și „stare de veghe *fără răspuns la stimuli*” – un aparent oximoron ce relevă deosebirea esențială dintre starea de veghe (vegetativă) și activare (afectivă) –, ceea ce Panksepp numește „intenționalitate” în citatul de mai sus[13]. Acesta este un alt motiv pentru care eu prefer termenul „activare” în loc de „stare de veghe” sau de „nivel” al conștiinței. „Activarea” poate include (și chiar sugerează) capacitatea de răspuns emoțional și intenționalitatea, care, după cum vedem din nou aici, se află în centrul comportamentului conștient[14]. Activarea afectivă permite *volițiunea*. Iată ce adaugă PAG la funcționarea automată, vegetativă.

Cum face acest lucru? PAG nu este parte din sistemul reticulat activator, deși se învecinează și e dens interconectată cu el[15]. Principala diferență dintre acești nuclei și PAG este direcția fluxului de informații dintre ele și prozencefal. Sistemul reticulat activator își exercită influența în principal *ascendent*, spre cortex, iar cortexul doar transmite semnale *descendent*, spre PAG[*2].

PAG este punctul final în care se adună toate circuitele cerebrale afective. Așadar, prozencefalul este activat de sistemul reticulat activator, iar PAG este activat (ca să zic aşa) de prozencefal. Ne-am putea gândi la sistemul reticulat activator și la PAG ca la originea și, respectiv, destinația activării prozencefalului.

În consecință, PAG este conceptualizată drept punctul terminus al rețelei „descendente” pentru afecte, pentru a fi pusă în contrast cu rețelele afective „ascendente” și „modulatoare” ale nucleilor care monitorizează corpul și ale sistemului reticulat activator din trunchiul cerebral[16]. Asta înseamnă că PAG este principalul centru de *output* al tuturor circuitelor afective, canalizând informațiile spre efectorii musculoscheletali și viscerali care generează „emoție propriu-zisă”[17]. (Aici citez direct din analize de specialitate, ca să mă asigur că ilustrez cu acuratețe modul în care înțelegem aceste funcții.) Rețeaua descendentă pentru afect este „implicată în acțiuni motoare specifice invocate în emoții, precum și în controlul frecvenței cardiace, respirației, vocalizării și comportamentului de împerechere”[18]. Rolul PAG în această rețea este de „*interfață* asupra stimulilor proeminenți între prozencefal și trunchiul cerebral inferior”[19]. În această privință, PAG este conceptualizată drept centru pentru „*echilibrarea sau transmiterea* informațiilor legate de predominanța supraviețuirii”[20]. Altfel spus, PAG funcționează „*orchestrând* diferite strategii de *coping* când este expusă la stresori externi”[21]. PAG „constituie un *punct de întâlnire* masiv al sistemelor neuronale care generează emotivitate”[22]. Astfel, ea joacă rolul central în „apărarea homeostatică a răspunsului individului, *integrând* informației aferente de la periferie și informații din centre superioare”[23].

Încerc să reformulez cât mai direct: *toate circuitele afective converg în PAG, care este principalul centru de output pentru sentimente și comportamente emoționale*. Din acest motiv, „stimularea electrică la intensități mai joase a acestei zone cerebrale va activa la animale o mai mare diversitate de acțiuni

emoționale coordonate decât stimularea oricărei alte regiuni cerebrale”[24].

Ca aspect, coloanele slab diferențiate ce constituie PAG înconjoară canalul central al trunchiului cerebral pe o lungime de 14 mm. Canalul central, prin care circulă un lichid incolor (lichidul cefalorahidian), este apeductul lui Sylvius – numit după anatomistul din secolul al XVII-lea care l-a descris pentru prima dată. Localizarea sa în centrul mezencefalului îi dă numele PAG: „substanța cenușie periapeductală” înseamnă, pur și simplu, „substanța cenușie care înconjoară canalul”[25]. Acest centru primitiv al creierului este cu adevărat ceea ce a descris Freud drept „straturile sale intime”. Se împarte în două grupuri de coloane funcționale[26]. Una dintre ele, cea din spate, este pentru „strategii de *coping*” *active* sau pentru comportamente defensive precum reacțiile de tip luptă-sau-fugă, tensiunea arterială crescută și reducerea durerii pe căi ce nu au legătură cu opioidele[27]. Aici se termină circuitele FRICII, FURIEI și PANICII/MÂHNIRII. Coloana din față este pentru strategii defensive/de *coping pasive* precum încremenirea cu hiporeactivitate, „comportamentul de bolnav” pe termen lung, tensiunea arterială scăzută și ameliorarea durerii cu opioide[28]. Circuitele DORINȚEI SEXUALE, GRIJII și CĂUTĂRII se termină în această coloană din față.

PAG este *calea comună finală* spre outputul afectiv. Prin urmare, ea trebuie, într-un cuvânt, „să aleagă” ce e de făcut în continuare, odată ce diferitele circuite afective și comportamentele condiționate aferente și-au adus contribuțiile la acțiune[29]. Ea trebuie să facă aceste alegeri prin evaluarea semnalelor de eroare reziduale ce îi sunt retransmise prin sistemele afective. Trebuie să le analizeze cererile concurente

în funcție de imperativele biologice supreme ale supraviețuirii și reproducerii, fiecare semnal de eroare comunicându-i nevoia componentă. Pe scurt, PAG trebuie *să stabilească prioritățile* pentru următoarea secvență de acțiune.

Dar acțiunile prioritare nu pot fi determinate exclusiv în funcție de nevoi. Nevoile trebuie să fie *contextualizate* nu numai reciproc, ci și de oportunitățile predominante.

Să luăm exemplul controlului respirator, descris mai sus. Să ne gândim mai întâi la contextul intern: dacă simt alarma de sufocare și sete în același timp, alarma de sufocare trebuie să fie prioritizată în fața setei, dar setea ar putea să aibă prioritate, de pildă, în fața tristeții. În al doilea rând, să ne gândim la contextul extern: dacă resimt o alarmă de sufocare, contextul îmi va spune dacă am nevoie – de exemplu – să elimin o obstrucție a căilor respiratorii sau să ies dintr-o încăpere plină cu dioxid de carbon. Contextul extern în care activez rămâne relevant pe măsură ce evenimentele se desfășoară, iar eu tot iau hotărâri privind ce să fac în continuare.

Să ne amintim definiția afectului din capitolul 5; ea include multe lucruri pe care oamenii nu le consideră de regulă a fi afective. Astfel, de exemplu, alarma de sufocare și setea nu sunt doar senzații corporale, ci, la fel ca anxietatea de separare, de pildă, transmit şi valori intrinsece – ceva bun și ceva rău din punct de vedere biologic. Durerea e un exemplu util pentru a face distincție între valența intrinsecă a afectului și contextul său exteroceptiv. *Senzația* neplăcută de durere e ceea ce o face afectivă, iar senzația somatică exteroceptivă transmite *localizarea* unui stimul dureros: „durere ce vine din mâna stângă”. Aceasta nu este o distincție filosofică – aspectele duale pot fi manipulate separat, de exemplu, prin stimularea PAG și,

respectiv, a cortexului parietal. Prin urmare, în anumite condiții clinice e posibil *să percepi* că te-ai înțepat cu o piuneză în mâna stângă fără *să simți* durere[30].

Pe acest fond, Björn Merker face o observație profundă[31]. Deși din punct de vedere anatomic PAG este localizată sub cortex, ea e cea mai importantă din punct de vedere funcțional. *După ce* cortexul și alte structuri ale prozencefalului și-au făcut treaba cognitivă – după ce și-au adus contribuția la acțiune –, *decizia finală* despre „ce să fac în continuare” se ia la nivelul mezencefalului. Asemenea decizii iau forma trăirilor afective, generate de PAG, care pot să aibă precedență asupra oricăror strategii cognitive formulate în timpul secvenței anterioare de acțiune. În orice moment, PAG alege afectul care va determina și va modula secvența următoare. De exemplu, în prima mea zi de lucru ca neuropsiholog, am decis *din punct de vedere cognitiv* să revin la paturile pacienților ca să le citesc fișele, dar ceea ce s-a întâmplat de fapt *din punct de vedere afectiv* este că am leșinat.

Desigur, viața nu constă din urgențe nesfârșite; cel mai adesea, va prevala activitatea implicită a CĂUTĂRII (interesul și implicarea de fond). Totuși, în situații de urgență, numeroasele semnale de eroare ce transmit nevoi și ajung în mezencefal nu pot fi percepute toate odată, deoarece nu poți *să faci* tot odată. Așadar, aceasta este esența funcției de prioritizare a PAG: care dintre aceste semnale de eroare sunt predominante, date fiind circumstanțele de moment? Care dintre problemele mele poate fi amânată (sau gestionată automat) și în cazul căreia trebuie să caut soluţia în funcţie de ceea ce simt?

Cerința că trebuie să ții cont într-un mod firesc de circumstanțele curente presupune un aparat suplimentar. PAG

își dă verdictul cu ajutorul unei structuri adiacente din mezencefal, numită *coliculii superiori* (vezi figura 6). Această structură se află chiar în spatele ei și este împărțită în mai multe straturi, fiecare oferind reprezentări simple derivate din simțurile corpului. Straturile mai profunde furnizează hărți motoare ale corpului, iar straturile superficiale sunt responsabile pentru cele spațio-senzoriale. Împreună, ele alcătuiesc o reprezentare extrem de comprimată și de integrată a lumii exteroceptive, provenind parțial din cortex, dar și din regiuni subcorticale senzorimotoare precum nervul optic (din nou, vezi figura 6). Astfel, coliculii superiori reprezintă într-o formă condensată starea corpului *obiectiv* (senzorial și motor) moment cu moment, cam așa cum PAG îi monitorizează starea *subiectivă* (nevoia). Merker numește această interfață afectivă/senzorială/motoare dintre PAG, coliculii superiori și zona locomotoare din mezencefal „triunghiul de decizie" al creierului[32]. Panksepp a numit-o SINELE primar, însăși sursa ființei noastre capabile de simţire[33].

Prin urmare, deciziile mezencefalului despre ce să faci în continuare se bazează pe feedback din partea circuitelor afective cerebrale *alături de* hărțile senzorimotoare, fiecare dintre acestea actualizându-ne diferitele aspecte ale „felului cum stau lucrurile acum". Să ne amintim de exemplul în care am conștientizat brusc nevoia de a urina la sfârșitul unei prelegeri de două ore. În acel moment, simțind ocazia, nevoia mea corporală a fost percepută și a devenit o *pulsiune* volițională. Pe scurt, deoarece triunghiul de decizie din mezencefal ia în considerare atât condițiile interne, cât și pe cele externe, el prioritizează opțiunile comportamentale nu numai pe baza nevoilor curente, ci și a oportunităților curente.

Stratul cel mai profund al coliculilor superiori constă într-o hartă ce controlează mișcările ochilor. Această hartă este în sine mai stabilă decât hărțile senzoriale superioare, întrucât celelalte hărți sunt calibrate în funcție de ea, stabilind astfel „punctul de vedere” unificat ce caracterizează experiența perceptivă subiectivă. Acesta este motivul pentru care ne percepem pe noi înșine în interiorul unei scene vizuale stabile indiferent cât de mult ni se mișcă ochii, căci se mișcă de aproximativ trei ori pe secundă. De asemenea, scena stabilizată sugerează că ceea ce percepem asta și este – o *scenă* –, o perspectivă construită pornind de la realitate, nu realitatea însăși. Tot din acest motiv ne percepem pe noi înșine ca trăind în mintea noastră[34]. Voi explica mai detaliat natura virtuală a percepției în capitolul 10.

După cum am văzut în cazul vederii oarbe, „harta bidimensională asemenea unui ecran” a lumii senzorimotoare, cum o numește Merker, este inconștientă la oameni[35]. Ea nu prea conține decât o reprezentare a direcției „deviației-țintă” – ținta fiind punctul asupra căruia se concentrează fiecare ciclu de acțiune –, producând privirea intenționată, atenția și orientarea acțiunii. Brian White o numește hartă „a predominanței” sau „a priorității”. Panksepp arată că acesta este modul în care „deviațiile noastre de la o stare de repaus ajung să fie reprezentate ca stări de disponibilitate pentru acțiune”[36]. Mai bine de atât nu cred că se poate formula.

Conștientizarea perceptivă a lumii din jurul nostru devine posibilă cu ajutorul cortexului activat corespunzător, care (spre deosebire de conștiința afectivă) este ceea ce le lipsește copiilor cu hidroanencefalie și animalelor ce au suferit decorticare cerebrală. Coliculii superiori furnizează mapări condensate de

tip aici-și-acum ale țintelor și acțiunilor potențiale, dar cortexul furnizează „reprezentări” detaliate pe care le folosim pentru a ghida fiecare secvență de acțiune pe măsură ce se desfășoară. Pe lângă aceste imagini foarte diferențiate, în prozencefalul subcortical există multe programe de acțiune *inconștiente* ce sunt numite „proceduri” și „răspunsuri” – nu imagini. (Să ne gândim, de exemplu, la tipurile automatizate de memorie pe care ne bazăm ca să mergem pe bicicletă sau ca să parcurgem traseul spre un loc familiar.) Acestea sunt codificate în principal în ganglionii bazali *subcorticali*, amigdală și cerebel. Amintirile nu sunt simple consemnări ale trecutului. Din punct de vedere biologic, ele sunt *despre* trecut, dar *pentru* viitor. În esență, toate sunt niște *predicții* menite să vină în întâmpinarea nevoilor noastre. Voi relua această idee importantă în capitolele următoare.

Tendințele motoare activate prin selecția mezencefal-afect deblochează reflexe și instincte simple – și asta e tot ce fac în cazul bebelușilor, al copiilor cu hidroanencefalie și al multor animale. Dar, după cum știm de-acum, asemenea comportamente automate intră sub un control individualizat pe parcursul dezvoltării, prin învățarea din experiență. Drept urmare, răspunsurile stereotipe sunt completate de un repertoriu mai flexibil de opțiuni. Secvența comportamentală ce apare cu fiecare ciclu de acțiune nou se desfășoară ascendent deasupra acestor niveluri ale controlului din prozencefal ce se extind progresiv, de la „răspunsuri” procedurale la „imagini mnezice” reprezentaționale. Așa apare ceea ce Merker numește „o lume tridimensională, panoramică, complet articulată, compusă din obiecte solide cu formă definită: lumea experienței noastre fenomenologice

familiare”[37].

Să ne amintim că memoria pe termen lung servește viitorului. Odată ce triunghiul de decizie din mezencefal a evaluat feedbackul comprimat ce vine din fiecare acțiune anterioară, el activează un proces de *feedforward* extins ce se desfășoară în sens invers, prin sistemele de memorie din prozencefal, generând un *context așteptat* pentru secvența motoare selectată. Acesta este produsul întregii noastre învățări. Cu alte cuvinte, când o nevoie ne scoate în lume, *nu redescoperim lumea de fiecare dată, cu fiecare ciclu nou*. Ea activează un set de predicții despre consecințele senzoriale probabile ale acțiunilor noastre, în baza experienței noastre din trecut de satisfacere a nevoii selectate în împrejurările predominante.

Apoi, acțiunea voluntară declanșează un proces de testare a *așteptărilor* noastre în comparație cu consecințele *reale* ale acțiunilor noastre. Această comparație generează un semnal de eroare, pe care îl folosim ca să ne reevaluăm așteptările pe parcurs și ca să ne ajustăm planurile de acțiune în mod corespunzător. La asta se referă comportamentul „voluntar”: să decizi ce să faci în condiții de incertitudine, în funcție de consecințele fiecărei acțiuni. Asta presupune consecințe resimțite atât afectiv, cât și perceptiv, motiv pentru care şi semnalele de eroare afective, și cele reziduale senzorimotoare converg în triunghiul de decizie din mezencefal. Îmi place cum numeşte Jakob Hohwy procesul psihic ce controlează comportamentul voluntar: „prezicerea prezentului”[38].

Ca să-ți faci o idee despre cum arată din punct de vedere clinic ceea ce spun, gândește-te la ce s-a întâmplat cu domnul S., inginerul electronist cu psihoză Korsakoff al cărui caz l-am

descris în capitolul 2. El se aștepta să vadă Johannesburgul când s-a uitat pe fereastra din cabinetul meu, lucru cauzat de tulburarea lui de memorie: nu își actualizase modelul predictiv astfel încât să corespundă cu evenimentele recente. Când peisajul londonez cu zăpadă pe care l-a văzut în realitate nu s-a potrivit cu așteptările lui, domnul S. nu și-a schimbat abordarea. A ignorat semnalul de eroare pe care îl primea și a rămas la predicția lui inițială, spunând: „Nu, *știu* că sunt în Johannesburg – dacă mănânci pizza nu înseamnă că ești în Italia”. Cu alte cuvinte, creierul său a modulat descendent semnalul de eroare. În consecință, domnul S. nu și-a ajustat programul de acțiune, fapt care nu avea cum să nu se termine într-un mod neplăcut – pentru că nu a reușit să învețe din experiență. Dacă familia și medicii nu s-ar fi îngrijit de toate nevoile sale, domnul S. cu siguranță ar fi murit[39].

Majoritatea oamenilor nu își dau seama că percepțiile noastre de tip acum-și-aici sunt ghidate permanent de predicții, generate în principal din memoria pe termen lung. Dar așa stau lucrurile. Din acest motiv, mult mai puțini neuroni propagă semnale de la organele senzoriale externe la sistemele mnezice interne decât invers[40]. De exemplu, raportul dintre conexiunile care intră și cele care ies în corpul geniculat lateral (ce transmite informație de la ochi la cortexul vizual și invers; vezi figura 6) este de aproximativ 1:10. Cea mai importantă contribuţie o au semnalele predictive care le *întâmpină* pe cele senzoriale ce sosesc de la periferie. Astfel se economisește un volum enorm de prelucrare a informațiilor și, prin urmare, travaliu metabolic. Având în vedere că creierul consumă cam 20% din rezervele noastre totale de energie, această eficiență e prețioasă. De ce să tratezi orice lucru ca și cum nu l-ai mai

întâlnit niciodată? În schimb, creierul propagă spre interior numai acea parte a informațiilor primite care *nu* se potrivește cu așteptările sale. Din acest motiv, astăzi percepția este descrisă uneori drept „fantezie” și „halucinație controlată”; începe cu un scenariu *așteptat* care este apoi ajustat astfel încât să se potrivească cu semnalul primit[41]. În acest sens, anatomiștii clasici au avut dreptate: procesarea corticală constă în principal din activarea unor „imagini mnezice”, rearanjate corespunzător astfel încât să prezică următorul ciclu de percepție și acțiune.

Afirmațiile mele sunt valabile în cazul tuturor vertebratelor, nu numai al oamenilor. Poate că sunt valabile chiar în cazul tuturor organismelor înzestrate cu creier sau cu sistem nervos. (Insectele, de exemplu, au structuri cerebrale care funcționează într-un mod similar cu sistemul nostru reticulat activator[42].) Știm deja că forma fundamentală a conștiinței este afectul, care ne permite să ne „simțim” calea de ieșire din situații neanticipate. Dar cum anume se transformă afectul endogen într-o exterocepție conștientă?

Panksepp sugerează că o punte evolutivă între aceste două aspecte ale experienței ar fi putut fi construită de „afectele senzoriale” (de exemplu, durerea, dezgustul și uimirea). Afectele senzoriale sunt în același timp senzații interioare *și* percepții exterioare; ele sunt în mod inerent percepții cu valențe condiţionate de senzații specifice. Așadar, de exemplu, durerea se simte diferit faţă de dezgust și răspundem la ele în mod diferit, retrăgându-ne sau vomitând, în funcție de cea pe care o resimțim. Conform lui Panksepp, pe această punte evolutivă conștiința s-a extins asupra percepției în general,

întrucât contextualizează afectul. Până la urmă, lumea exterioară dobândește valoare doar pentru noi deoarece este locul unde trebuie să ne satisfacem nevoile; ea capătă valoare *în raport cu* afectul.

Așadar, răspunsul meu la întrebarea de ce resimțim percepția, acțiunea și cogniția este că *le resimțim pentru că ele contextualizează afectul*. E ca și cum experiența noastră perceptivă spune: „Simt așa *despre asta*”[43]. Cum s-ar spune, percepția este incertitudine aplicată, afirmație ce justifică opinia lui Merker conform căreia conștiința afectivă și cea perceptivă utilizează o „monedă comună” – ele sunt *tipuri* diferite de sentimente, dar sunt sentimente[44]. Aici vorbesc numai despre percepția *conștientă*. Iar de aici nu mai e decât un pas mic până la a deduce că cele cinci modalități de conștiință perceptivă ale omului (văzul, auzul, pipăitul, gustul și mirosul)[45] au evoluat pentru a descrie diferitele categorii de informații exterioare înregistrate de organele noastre senzoriale[46], așa cum cele șapte tipuri de conștiință afectivă discutate în capitolul anterior descriu diferitele categorii de nevoi emoționale la mamifere.

Dar este important să ne amintim și că percepția corticală este *apercepție*. În conștiință nu apar semnalele senzoriale brute transmise de la periferie, ci *inferențe* predictive derivate din urmele mnezice ale unor asemenea semnale și ale consecințelor lor. Faptul că abilitatea de a simți se întinde, ca să zic aşa, peste inferențele corticale stabilizate creând ceea ce eu numesc „corpuri solide psihice”[47] – lumea exterioară așa cum se manifestă ea în percepție – explică parțial calitățile fenomenologice diferite ale afectului versus percepția conștientă. Voi reveni asupra acestor teme în capitolul 10.

Modelul interior al creierului este harta pe care o folosim pentru a naviga prin lume – chiar pentru *a genera* o lume preconizată. Dar nu ne putem lua drept bune toate predicțiile. De fapt, modelul interior generează două aspecte ale „contextului preconizat”: pe de o parte, avem conținutul propriu-zis al predicțiilor noastre, iar pe de altă parte, *nivelul de încredere* în acuratețea lor. Întrucât toate predicțiile sunt probabilistice, trebuie codificat și gradul de *incertitudine preconizată* aferentă. Predicțiile sunt furnizate de rețelele memoriei pe termen lung din prozencefal, care filtrează prezentul prin prisma trecutului. Dar a doua dimensiune – ajustarea nivelurilor de încredere – este esența activității desfășurate prin activarea modulatoare.

Asta s-a dereglat în cazul domnului S.: a pus prea multă bază în predicțiile lui. Cu alte cuvinte, din cauza afecțiunii sale, nu a acordat suficientă importanță semnalelor de eroare (le-a acordat prea puțină *atenție*)[48]. Cum e posibil acest lucru? Într-un fel, am analizat deja acest lucru din perspectiva cealaltă. Modularea semnalelor neuronale înseamnă pur și simplu ajustarea în sus sau în jos a intensității lor. E ceea ce face triunghiul de decizie din mezencefal atunci când selectează ce semnale să intensifice – *modulează* intensitatea semnalului prin intermediul sistemului reticulat activator. Intensitatea inițială a semnalului reprezintă contextul *preconizat*. Totuși, pe măsură ce se dezvăluie contextul real (resimțit), intensitățile semnalelor trebuie ajustate. De aici decurge că încrederea sporită într-un semnal de eroare implică cu necesitate o încredere scăzută în predicția ce a dus la eroare. Dacă ți se declanșează alarma de fum, dar nu pare să ardă nimic, de vină sunt fie alarma, fie

aparențele și, cumva, trebuie dat un verdict în ce privește această neconcordanță. Asta se întâmplă și în triunghiul de decizie din mezencefal: se evaluează cererile concurente și se declară un câștigător. Rezultatul acestui proces este transmis spre prozencefal de sistemul reticulat activator, care acționează în funcție de preconizările tale și apoi – pe măsură ce acțiunea selectată se desfășoară – eliberează nori de molecule neuromodulatoare ca să rafineze semnalele din rețelele memoriei tale pe termen lung, reglând pozitiv canalele prozencefalului în care sunt depozitate unele predicții și negativ altele. La rândul său, acest lucru duce la învățarea din experiență. Astfel, continuăm să ne îmbunătățim modelul generativ al lumii, prin încercare și eroare.

Tot ceea ce facem pe tărâmul incertitudinii este ghidat de aceste niveluri fluctuante ale încrederii. Credem că știm ce se va întâmpla dacă acționăm într-un anumit fel, dar oare chiar știm? Dacă convingerea ne scade sub un anumit prag, atunci nu acționăm sau schimbăm abordarea. În sfera exteroceptivă, treaba merge bine când se desfășoară conform așteptărilor și prost când predomină incertitudinea. În consecință, *simțim* că treaba merge bine sau prost: creșterea încrederii (într-o predicție) este bună, iar scăderea încrederii este rea. Deci încercăm să minimizăm incertitudinea în așteptările noastre.

După cum am văzut mai sus, percepția și acțiunea sunt un proces continuu de *testare a ipotezelor*, în care creierul încearcă permanent să elimine semnalele de eroare și să-și confirme ipotezele. (Știința experimentală în sine este doar o versiune sistematizată a acestui proces de zi cu zi, după cum am văzut în capitolul 1: „Dacă ipoteza X este corectă, atunci, când fac Z, ar trebui să se întâmple Y”.) Cu cât ți se confirmă

ipotezele, cu atât ești mai încrezător și cu atât mai puțin activat – mai puțin conștient – ai nevoie să fii. Îți poți automatiza secvențele de acțiune, trecând în starea implicită. Dar, dacă ajungi într-o situație neprevăzută – pe care modelul tău predictiv nu pare să o clarifice într-un mod demn de încredere –, consecințele acțiunilor tale devin predominante. Ieși de pe pilot automat și devii hiperconștient: triunghiul de decizie îți ajustează cu grijă predicțiile pe măsură ce cauți, în funcţie de ceea ce simți, o cale de a evita consecințele acțiunilor tale și faci noi alegeri.

Astfel, simțirea, atât afectivă, cât și cognitivă, rămâne factorul comun în orice formă de conștiință. Funcția simțirii este să evalueze succesul sau eșecul programelor tale de acțiune și al contextelor aferente acestora. Dar nu toate eșecurile sunt la fel de semnificative. Sentimentele cele mai puternice au legătură cu incertitudinile importante pentru supraviețuire, și nu cu cele atașate de posibilități relativ minore. Predicţiile proaste precum „Mașina pe care o urmăresc o va lua acum la stânga” pot fi gestionate la niveluri periferice din ierarhia predictivă, unde este tolerat un grad mai mare de eroare, și, prin aceasta, determină schimbări în ceea ce numim atenție, nu afect (deși atenția este modulată în egală măsură de trunchiul cerebral reticulat). Totuși, dacă deodată crește probabilitatea ca mașina pe care o urmărești să te implice într-un accident, probabil că eroarea ta va activa un răspuns afectiv. Atenția poate fi şi direcționată, și captată. În al doilea caz e însoțită de sentimente de uluire sau spaimă, ce rescriu din mers predicțiile eșuate, astfel încât modelul lumii să nu îți mai facă aceeași surpriză neplăcută.

Iată cum funcționează conștiința în creier, la nivel fiziologic.

Poate că nu pare cu totul satisfăcător. De pildă, nu prea mi-am bătut capul cu abordările filosofice ale problemei minte/corp. Nu m-am ocupat de relația dintre ceea ce unii filosofi numesc „aspectele ei duale”: de ce fiziologia obiectivă a conștiinței este însoțită de o senzație fenomenologică subiectivă? Atunci când oamenii de știință observă o asemenea corelație regulată, între oricare două seturi de date, ei caută o singură cauză fundamentală. Vor *să explice de ce* fenomenele se întâmplă în același timp. Așadar, ca să înțelegem legăturile regulate dintre aspectele fiziologice și cele psihologice ale conștiinței, trebuie să pătrundem mai în profunzime. Trebuie să depășim constrângerile disciplinare ale fiziologiei și psihologiei. De obicei, noi, oamenii de știință, facem acest lucru apelând nu la metafizică, ci la fizică[49]. În acest caz, răspunsurile de care avem nevoie trebuie căutate în fizica entropiei[*3].

[*1] De remarcat că unele molecule neuromodulatoare pot acționa *și* ca neurotransmițători.

[*2] În toată cartea, folosesc cuvintele „deasupra”, „dedesubt”, „mai sus”, „mai jos”, „ascendent”, „descendent” etc. nu ca judecăți de valoare, ci ca localizatori anatomici. Spre deosebire de alte organe, creierul este ierarhic *din punct de vedere structural*. E stratificat oarecum asemenea unui sit arheologic, nivelurile mai vechi fiind acoperite de cele mai noi. De unde și titlul acestei cărți: Izvorul *ascuns*. Centrul cel mai profund al trunchiului cerebral conține structurile cele mai vechi, din punctul de vedere al evoluției, iar nivelurile superioare ale cortexului le conțin pe cele mai recente. Asta nu înseamnă că structurile inferioare (și mai vechi) sunt mai puțin importante decât cele superioare (mai noi).

Dimpotrivă, din punct de vedere funcțional, structurile cele mai de sus ale prozencefalului sunt pur și simplu *dezvoltări* ale structurilor celor mai de jos din trunchiul cerebral.

*3 Cititorii care nu au urmărit notele de final ar putea consulta acum anexa de la p. 313. Deși e oarecum tehnică, ea face o trecere utilă spre capitolul următor.

7

Principiul energiei libere

Un grup de oameni de știință sunt într-un laborator, urmărind un monitor mare de calculator. Puncte și pete se rotesc pe ecran. Se pot distinge diferite culori: albastru, roșu, violet și altele. Și punctele par să aibă mărimi diferite, dar mișcările lor haotice nu au un tipar. Se vălătucesc și se întrepătrund ca niște nori de gaz, umplând la întâmplare spațiul virtual. Cu toate acestea, un ceas digital marchează trecerea timpului în secunde, iar două axe în partea de jos a ecranului, în stânga, identifică localizarea spațială a fiecărui punct. Asta sugerează că se întâmplă ceva măsurabil. Dar cum se pot cuantifica mișcările în acest haos?

Îl întrebi pe unul dintre fizicienii din încăpere la ce ne uităm, iar el spune, fără ca răspunsul să te ajute, că este un proces „stocastic”. Stocastic înseamnă aleatoriu. Treptat, observi că punctele mai mici par să fie mai lente decât cele mai mari; par să danseze pe o melodie ușor diferită. La indicaţiile unui neurolog din încăpere, informaticianul scrie ceva la tastatură, iar vârtejurile de pe ecran încremenesc brusc. Informaticianul salvează atent datele. Apoi, tot la cererea neurologului, tastează o mulțime de cifre care schimbă valorile

într-un set de ecuații care acum apar pe un ecran mic de lângă cel mare la care privești. Îţi spune că ajustează „interacțiunile locale ale subsistemelor”.

Din nou, particulele ca niște nori apar în colțul din stânga jos al ecranului și încep să se rotească. De data asta, după o scurtă perioadă de haos, e mai ușor să discerni un tipar. Punctele colorate fac valuri în principal spre exterior la început, apoi, treptat, se adună în mase și converg spontan spre centrul ecranului, formând o masă mai degrabă circulară. Cu puțină imaginație, ar putea fi un stol compact de păsări (poate grauri) care zboară coordonat. Constant, mișcarea particulelor devine mai ordonată. Se înghiontesc ca niște soldați ocupându-și pozițiile stabilite pentru o paradă, până ce începe să se întrevadă o structură clară: patru straturi concentrice. În centru sunt punctele albastru-închis. Acestea sunt înconjurate de cele roșii, care, la rândul lor, sunt înconjurate de cele violet. Punctele albastru-deschis formează un fel de limită exterioară. Toate acestea se petrec pe un fond de puncte negre mai mici care par să continue să se miște haotic. Neurologul pare mulțumit. Îi cere informaticianului să oprească mișcările de pe ecran. Ceasul arată că au trecut 1.278 de secunde.

Ne aflăm la Centrul Wellcome pentru Neuroimagistică Umană din Queen Square, Londra, unde Karl Friston este directorul științific. Te-a invitat să urmărești un experiment interesant: o simulare a interacțiunilor pe distanțe scurte ce apar între diferite subsisteme fizice atunci când sunt supuse unor forțe diverse.

Regulile ce guvernează aceste particule virtuale sunt la fel de variate ca şi cele care guvernează comportamentul moleculelor și atomilor reali: tendințe neintenționate (dar nu la

întâmplare) de atracție și de respingere. Evident, asemenea interacțiuni generează *ordine din haos*. Se consideră că astfel de ordonări spontane au avut loc atunci când a apărut viață din supa primordială. Totuși, experimentul se desfășoară într-un centru de neuroștiințe cognitive, vizavi de Spitalul Național de Neurologie și Neurochirurgie, și nu se supără nimeni dacă te întrebi ce legătură au cu creierul aceste particule virtuale.

Conform explicațiilor lui Friston, sistemele biologice precum celulele trebuie să fi apărut prin intermediul unor variante complexe ale aceluiași proces care a format sisteme mai simple „cu autoorganizare", cum sunt cristalele dintr-un lichid, pentru că au în comun un anumit mecanism. Acest mecanism, spune Friston, este „minimizarea energiei libere" (pe care îl voi explica puţin mai jos). Toate sistemele cu autoorganizare, inclusiv tu și eu, au în comun o sarcină fundamentală: să-și continue existența. Friston crede că realizăm această sarcină prin minimizarea energiei noastre libere. Cristalele, celulele și creierii – spune el – nu sunt decât manifestări tot mai complexe ale acestui mecanism fundamental ce se autoconservă[1]. Într-adevăr, atât de multe aspecte a ceea ce considerăm a fi viață psihică apar în zorii organizării biologice, încât contribuția creierului propriu-zis poate să pară destul de subtilă. Dar, atât timp cât nu pierdem din vedere conceptul de energie liberă, totul – chiar totul – va deveni clar.

Karl Friston este ultimul dintre marii oameni de știință care m-au influențat din punct de vedere profesional. Îl consider un geniu și (în mod obiectiv) specialistul în neuroștiințe cel mai influent din lume la ora actuală. Influența se măsoară prin „indicele Hirsch", ce măsoară impactul lucrărilor pe care le-ai

publicat[2]. Ca regulă generală, dacă indicele tău Hirsch este mai mare decât numărul anilor de când ți-ai obținut doctoratul, atunci stai bine. Indicele Hirsch al lui Friston este 235, mai mare decât al oricărui alt specialist în neuroștiințe[3]. A devenit celebru cu „reprezentarea statistică parametrică”, ce a permis analiza neuroimagisticii funcționale atât de răspândită astăzi. Totuși, activitatea sa în ce privește „codificarea predictivă” și principiul energiei libere i-au adus mai multă recunoaștere.

În pofida reputației extraordinare a lui Friston, timp de mulți ani nu prea m-am interesat de activitatea sa. Apoi, în 2010, a publicat o lucrare împreună cu un tânăr psihofarmacolog pe nume Robin Carhart-Harris, de care știam că e interesat de neuropsihanaliză. Lucrarea acestor doi autori argumenta că noțiunea de energie pulsională (adică „energie psihică”) a lui Freud corespunde principiul energiei libere[4]. După cum am explicat mai sus, Freud a admis imediat că „aici s-a petrecut ceva complet insesizabil pentru noi”, cu referire la modul în care nevoile corpului ar putea deveni energie psihică. A scris și că această energie se pretează la creștere, reducere, dislocare și descărcare, având astfel toate caracteristicile unei cantități, „deși nu avem niciun mijloc prin care să o măsurăm”[5]. Ţinând cont că intenția inițială (abandonată) a lui Freud fusese de a „prezenta procesele psihice ca pe nişte stări determinate cantitativ”, lucrarea lui Carhart-Harris și Friston a fost ca o lovitură de trăsnet pentru mine. Dacă energia mintală chiar este izomorfă cu schimbările energiei libere termodinamice, atunci – m-am gândit eu – trebuie să fie şi măsurabilă, și reductibilă la legile fizicii.

Prin urmare, am studiat lucrările anterioare ale lui Friston și l-am căutat. Ne-am întâlnit de mai multe ori în anii ce au

urmat, la Londra și la Frankfurt. Principalul subiect al conversațiilor noastre a fost rolul afectului în viața psihică. Deoarece activitatea lui Friston din perioada aceea era legată preponderent, ca în cazul mai tuturor specialiștilor, de cortex, mecanismele predictive pe care le descoperise vizau aproape exclusiv *cogniția*. Din acest motiv, de exemplu, un celebru articol al său ce arată că codificarea predictivă explică modul în care neuronii comunică între ei are titlul „A theory of *cortical* responses”[6]. Trebuie să mărturisesc însă că nu înţelesesem pe deplin multe dintre articolele sale mai tehnice.

În 2017, Friston a fost invitat ca vorbitor principal la congresul nostru anual de neuropsihanaliză (ce s-a ținut în acel an în vechiul University College Hospital, Londra, având ca temă principală codificarea predictivă). Ca să-mi formulez observaţiile finale fără să mă fac de râs, trebuia să stăpânesc chestiunile de fizică. Prin urmare, printre multe alte lucrări ale lui Friston, i-am recitit un articol foarte tehnic apărut în una dintre revistele publicate de Royal Society. Se numea „Life as we know it”[7]. Cu mare efort, l-am înțeles corect pentru prima dată. Acest articol nu își propunea nimic altceva decât să reducă la niște ecuații matematice legile fundamentale ce guvernează *intenționalitatea*.

Implicațiile erau electrizante. Mi s-a părut că aceste ecuații îmi puteau oferi pasul înainte pe care îl căutam. Așa că, imediat după discuțiile noastre științifice de la congresul din 2017, i-am scris propunându-i să ne punem ideile laolaltă, ca să încercăm să încorporăm conștiința în principiul energiei libere. Spre încântarea mea, Friston mi-a acceptat propunerea și am început să colaborăm la un articol în care am schițat ceea ce a devenit opinia noastră comună[8].

Legătura dintre activitatea lui Friston și cercetările mele este *homeostaza*. Am explicat mai sus că trebuie să ne menținem între limitele noastre viabile din punct de vedere fiziologic. Să folosim termoreglarea ca exemplu. Nu te poți mulțumi cu orice temperatură corporală: trebuie să te menții în intervalul limitat 36,5-37,5ºC. Dacă te încălzești semnificativ mai mult, mori; dacă te răcești semnificativ mai mult, mori. Nu poți permite temperaturii corporale să devină egală cu cea a mediului ambiant, cum face apa fierbinte când o adaugi într-o cadă cu apă rece. Apa fierbinte care curge în cadă nu rămâne separată de cea rece, într-o globulă mare, sub robinet. Însă *tu* da – și aşa şi trebuie, ca să rămâi în viață –, iar asta presupune *travaliu*. Pacienții comatoși nu pot realiza un asemenea travaliu, așa că mor din cauze precum hipertermia; literalmente, se supraîncălzesc[9]. Același lucru e valabil şi în cazul reglării gazelor sangvine, al echilibrului fluidelor și al celui energetic și al multor altor procese fiziologice. E valabil chiar și în cazul nevoilor *emoționale*, care, după cum am văzut în capitolul 5, nu sunt mai puțin „biologice” decât cele corporale. Şi menținerea în limitele viabile ale emoțiilor noastre presupune travaliu: ca să rămânem aproape de persoanele care au grijă de noi, ca să scăpăm de prădători, ca să îndepărtăm obstacole și așa mai departe. Dincolo de un anumit nivel de predictibilitate, travaliul necesar ca să facem aceste lucruri este reglat de sentimente.

Mecanismul pe care tocmai l-am descris este o formă extinsă de homeostază și nu e complicat (vezi figura 12).

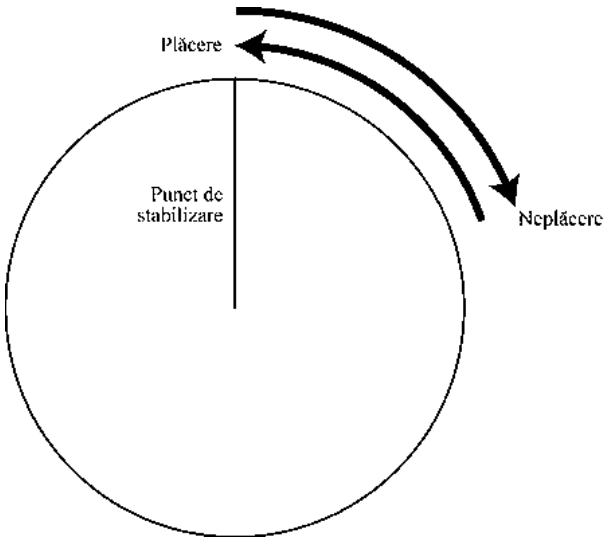


Figura 12. Homeostaza sentimentului. Punctul de stabilizare reprezintă limitele viabile ale sistemului.

Fiecare homeostat constă din doar trei componente: un *receptor* (ce măsoară temperatura, în exemplul meu), un *centru de control* (ce determină modul în care se menține temperatura în limitele viabile: 36,5-37,5ºC în exemplul meu) și un *efector* (care face munca necesară pentru a te readuce în limitele respective când le depășești). Pentru că mecanismul homeostazei este atât de simplu, poate fi redus la legi ale fizicii. Despre asta este vorba în articolul lui Friston: legile fundamentale ce guvernează „viața așa cum o știm”.

M-am entuziasmat când am realizat că aceste legi – extinse astfel încât să includă forma mai predictibilă de homeostază ce stă la baza afectului – ar putea să explice și conștiința: și nu numai elementele observabile exterioare ale comportamentului conștient (*fiziologia* obiectivă a triunghiului de decizie din mezencefal și a sistemului reticulat activator), ci și cele observate în interior. Aceasta promitea să explice *sentimentele* subiective care ghidează deciziile.

În acest punct, drumul meu şi cel al lui Damasio s-au despărțit. Poate mai mult decât ceilalți specialiști în neuroștiințe, el atrăsese atenția asupra faptului că afectul (și, prin urmare, conștiința) este la bază o formă de homeostază. Dar a respins inferența mea suplimentară: în cazul în care conștiința este homeostatică, iar homeostaza este reductibilă la legile fizicii, atunci și fenomenele conștiinței pot fi reduse la legi din fizică[10]. Mecanismul conștiinței, nu mai puțin decât cel al mișcării corpurilor cerești și a orice altceva din natură, *trebuie* să poată fi explicat în baza unor legi și, prin urmare, să fie oarecum predictibil, chiar dacă doar la un nivel probabilistic. Când Damasio a citit o variantă preliminară a articolului pe care l-am scris împreună cu Friston, am avut o discuție tensionată în biroul lui. Nu putea să înțeleagă de ce încercam să reduc conștiința la ceea ce el numea „algoritmi”. Era cazul clasic al specialistului ce refuză implicațiile unei idei pe care el însuși a lansat-o în lume – ca răspunsul lui Einstein la mecanica cuantică: „Dumnezeu nu joacă zaruri cu universul”[11].

I-am expus pe scurt lui Damasio argumentaţia pe care o voi explica în cele ce urmează[12]. Totuși, ca să explic cum trebuie, este necesar să prezint fizica pe care am fost nevoit să o stăpânesc ca să înțeleg principiul energiei libere. Trage adânc aer în piept.

Esența homeostazei este că organismele vii trebuie să ocupe un număr limitat de stări fizice: stările lor viabile ori stările lor valoroase sau preferate ori ceea ce Friston numește (referindu-se la toate acestea) stările lor „preconizate”. Nu ne putem permite să ne risipim în toate stările posibile. S-a dovedit că acest imperativ biologic are o legătură profundă cu unul dintre

conceptele explicative fundamentale din fizică, şi anume *entropia*. Majoritatea oamenilor înțeleg într-un mod intuitiv ce este „entropia”. O concep ca pe o tendință naturală spre dezordine, disipare, dezintegrare și altele asemenea. Legile entropiei sunt ceea ce face gheața să se topească, bateriile să se descarce, bilele de biliard să se oprească, iar apa fierbinte să se amestece cu cea rece.

Homeostaza merge în direcția opusă. *Se opune* entropiei. Se asigură că ocupi un număr limitat de stări posibile. În acest fel, îți menține temperatura necesară și te ține viu – și nu te lasă să te dezintegrezi. Ființele vii *nu au cum să nu* se opună unuia dintre principiile fundamentale ale fizicii: a doua lege a termodinamicii.

Prima lege a termodinamicii se referă la conservarea energiei[13]. Conform acesteia, energia nu poate fi creată sau distrusă, ci poate fi doar transformată dintr-un tip într-altul și poate trece dintr-un loc într-altul. (Mai știm, datorită lui Einstein, și că poate fi transformată în materie.)

Conform celei de-a doua legi, procesele naturale sunt întotdeauna ireversibile[14]. Astfel, apa fierbinte din cadă amestecată cu apa rece nu poate fi separată ulterior. De asemenea, energia din căldură nu poate fi pusă la loc în cărbunele ars care a produs-o și nici energia pierdută pe parcurs. Acest lucru se datorează entropiei, care, în consecință, *crește mereu* la scara universului[15]. De fapt, s-ar putea ca entropia să fie baza fizică a faptului că însuși timpul pare că are o direcție și o curgere.

În termodinamică, există două stări ale energiei: exergică și nonexergică. „Exergicitatea” energiei este definită prin capacitatea ei de a efectua lucru mecanic. De exemplu, energia

dintr-o bucată de cărbune poate fi arsă pentru a produce căldură[16], care poate fierbe apă pentru a produce abur, care poate pune în mișcare un motor. Dar în fiecare etapă a acestui proces se pierde o anumită cantitate de energie. Adică nu o poți folosi niciodată în întregime într-un mod util pentru lucru mecanic.

Coroborând aceste fapte, aflăm următoarele: pe măsură ce exergia dintr-un sistem se epuizează, entropia crește. În consecință, capacitatea sistemului de a efectua lucru mecanic scade întotdeauna. Prin urmare, entropia este asociată cu pierderea de exergie, din cauză că energia nu mai este disponibilă pentru a efectua lucru mecanic. Legea a doua este o afirmare a faptului ineluctabil că o parte din energie se pierde în lucru mecanic util în timpul oricărui proces natural[17].

Am menționat o anumită înțelegere intuitivă a „entropiei” puțin mai sus, dar definiția formală, tehnică a entropiei se referă la numărul de stări distincte pe care le poate ocupa un sistem[18]. Entropia este determinată de numărul de stări microscopice posibile care ar duce la aceeași stare macroscopică. Mai simplu spus: *cu cât stările posibile sunt mai puține, cu atât e mai scăzută entropia.*

Homeostaza stabilește limite ale tipurilor de stări macroscopice pe care le poate ocupa un sistem ca mine sau ca tine. Amintește-ți că homeostaza ne menține în viață efectuând un lucru mecanic; prin urmare, dacă entropia presupune pierderea capacității de a efectua lucru mecanic, ea e „rea” pentru noi, ca sisteme biologice. Funcția elementară a ființelor vii este să se opună entropiei.

Exemplul utilizat în mod convențional de fizicieni pentru a

ilustra aceste concepte implică un gaz comprimat introdus într-o cameră mai mare, goală, printr-un orificiu mic. Pe măsură ce moleculele se mișcă aleatoriu, ele explorează camera, răspândindu-se pentru a ocupa spațiul disponibil. Cu cât trece timpul, cu atât fiecare moleculă se poate găsi în mai multe locuri. Singura modalitate prin care poate fi inversat acest proces – singura modalitate prin care gazul poate fi readus în recipientul său inițial – presupune lucru mecanic.

Gândește-te la entropie prin raportare la numărul de locuri în care s-ar putea afla fiecare moleculă la un moment dat. Aceasta se dovedește a fi o afirmație despre *probabilitate*. Probabilitatea statistică a fiecărei molecule de a ocupa o anumită poziție se reduce pe măsură ce crește entropia: pe măsură ce gazul se dilată, poziția fiecărei molecule devine mai puțin predictibilă. Creșterea entropiei înseamnă *reducerea predictibilității.*

Acest lucru este important deoarece, spre deosebire de celelalte legi ale termodinamicii, legile probabilității se aplică tuturor lucrurilor, nu numai celor materiale. Așa cum entropia unui gaz dintr-o cameră poate fi definită sub aspect probabilistic, tot astfel și entropia poate fi asociată cu un proces psihologic de luare a deciziilor. În ambele cazuri, entropia crește odată cu caracterul aleatoriu al rezultatelor posibile. „Entropia” asociată cu gazele ce se dilată și cea asociată cu opțiunile al căror număr crește *sunt unul şi același lucru*. Nu toate lucrurile care există în natură sunt vizibile și tangibile, dar toate se supun legilor probabilității. Acesta este motivul pentru care probabilitatea tulbură fizica modernă în esența ei, unde materia nu mai este considerată un concept fundamental, iar particulele clasice dispar[19].

În exemplul fizic al entropiei, când un gaz a fost inițial comprimat în containerul său, iar moleculele sale erau înghesuite, pentru a descrie locul fiecărei molecule au fost necesari mai puțini *biți de informație* decât după ce gazul a fost eliberat pentru a umple spațiul disponibil în camera mare. În informatică, o „cifră binară” – pe scurt, „bit” – este unitatea elementară de informație. Un bit poate să aibă una din două valori opuse – de exemplu, da versus nu, pornit versus oprit, pozitiv versus negativ. De obicei, aceste stări sunt reprezentate ca 1 versus 0[20].

Inițial, sunt necesari mai puțini biți de informație pentru a descrie gazul, dar, pe măsură ce acesta se dilată, numărul de stări posibile pe care le poate ocupa fiecare moleculă crește. Astfel, entropia (măsurată de obicei din punct de vedere fizic în termeni de căldură schimbată și de temperatură) poate fi măsurată și în biți: cu cât e necesară mai multă informație pentru a descrie microstarea unui sistem (adică starea fiecărei molecule), cu atât crește entropia termodinamică. Sau, în termeni și mai simpli: *cu cât e mai mare numărul de întrebări cu răspunsuri de tip da/nu la care trebuie să se răspundă pentru a descrie un sistem, cu atât mai mare este entropia lui*. Astfel, când microstările unui sistem au valori de probabilitate scăzute, fiecare măsurare a sistemului conține mai multă informație decât atunci când ar avea o probabilitate mare – fiindcă ar trebui să se răspundă la mai multe întrebări binare pentru a descrie totalitatea stării sistemului.

Entropia este minimă când răspunsul la fiecare întrebare implicând răspunsuri de tip da/nu este cu totul predictibil, adică atunci când nu se află nimic și nu se obține nicio

informație. Conținutul informațional al datului cu banul este un bit, deoarece șansele să iasă cap sunt de 50%, dar obții zero informație când foloseşti un ban cu două capete, deoarece șansele să iasă capul sunt de 100%. Nu furnizează nicio informație pentru că răspunsul este în întregime predictibil. Entropia măsoară cantitatea medie de informație pe care o obții prin măsurarea unui sistem de mai multe ori. Astfel, entropia unei serii de măsurători este *informația sa medie*, *incertitudinea sa* medie.

Ca să ne dăm seama de importanța acestui lucru pentru o perspectivă neuroștiințifică asupra conștiinței, să ne amintim că tiparele cu unde lente sincronizate dintr-o electroencefalogramă (EEG) sunt mai ordonate (mai predictibile) decât cele rapide desincronizate (haotice). Prin urmare, tiparele „cu activare scăzută” conţin mai puțină informație decât cele „cu activare mare” (vezi figura 10). *Cele cu activare mai mare conțin mai multă incertitudine*[21]. Astfel, valorile entropiei din EEG sunt mai mari la pacienții aflaţi într-o stare minimă de conștiență decât la cei aflați în stare vegetativă[22]. Iar asta este explicabil: activitatea corticală din creierul conștient comunică mai multă informație decât în timpul somnului profund. Iar acum vine partea ciudată: dacă mai multă informație înseamnă mai multă incertitudine și, prin urmare, mai multă entropie, atunci – întrucât ființele vii trebuie să se opună entropiei – activitatea în stare de veghe este mai puțin dezirabilă, din punct de vedere biologic, decât somnul profund[23]. Știu că această afirmație este contrară logicii, dar va deveni mai ușor de înțeles pe măsură ce avansăm cu explicațiile[24].

Relația dintre entropie și informație a fost formalizată într-o

ecuație celebră de Claude Shannon, inginer electronist și matematician. Cu această descoperire, Shannon a încorporat „informația” în fizică, în cadrul căreia a devenit de atunci un concept fundamental, mai ales în mecanica cuantică[25]. Pornind de la ideile lui Shannon, fizicianul Edwin Thompson Jaynes a argumentat că entropia termodinamică trebuie considerată o *aplicație* a entropiei informaționale[26]. Prin urmare, definiția lui Shannon este și mai importantă decât cea din termodinamică: o definiție abstractă a entropiei în termeni de dinamică a informației este mai aplicabilă la un nivel general decât una concretă, în termeni de dinamică a căldurii. Așadar, legile termodinamicii pot fi considerate *un caz special al legilor mai profunde ale probabilității*. Acest lucru este important deoarece legile termodinamicii se aplică numai sistemelor materiale (tangibile, vizibile), precum creierul, iar legile informației se aplică și sistemelor imateriale (intangibile, invizibile), precum psihicul.

Dar probabilitatea nu este chiar totuna cu informația. Din perspectiva lui Shannon, informația presupune un factor suplimentar, *comunicarea*. De unde și titlul articolului său fundamental care a consacrat cibernetica: „Teoria matematică a comunicării”[27]. Spre deosebire de probabilități – care există în sine –, comunicarea presupune atât o *sursă* a informației, cât și un *receptor* al informației. (Nu e neapărat necesar ca comunicatorul să fie o persoană. Ar putea fi o carte, de pildă, sau orice sistem care deține informație din care receptorul poate să învețe.)

Acest lucru pune probleme mari oricărei ipoteze teoretice conform căreia conștiința *pur și simplu este* informație[28], apărând întrebarea: care este sursa și care este receptorul

informației – integrată sau de alt tip? Acesta este motivul pentru care eu nu sunt mulțumit de modelele fluxului de informații folosite de cognitiviști. Ei evită întrebarea „Unde este *subiectul*, receptorul?”. În acest fel, ca să-l parafrazez pe Oliver Sacks, psihicul este exclus din științele cognitive.

Această omitere a subiectului ce are experienţe duce la o întrebare și mai importantă totuși – poate cea mai importantă dintre cele cu care se confruntă astăzi științele cognitive: înainte de toate, în lipsa unui observator, cum și de ce se întâmplă prelucrarea informației (adică formularea unor întrebări, apoi a unor răspunsuri la ele)?

Descoperirea de către Shannon a informației-ca-entropie l-a determinat pe fizicianul John Wheeler să propună o interpretare „participativă” a universului[29]. Conform lui Wheeler, lucrurile apar în forma în care apar (cu alte cuvinte, ca fenomene observabile) numai ca răspuns la întrebările pe care le punem. Fenomenele ca atare există numai în ochii celui care le privește, ai unui observator participant, ai celui care pune întrebări. În celebra formulare a lui Wheeler, „cevaurile apar din biți” (unde „cevaurile” sunt *lucruri* observabile)[30]. Așadar, informația este „fizică” nu numai pentru că e implicată în legile fizicii, ci și pentru că e baza tuturor *fenomenelor observabile*. Așa devin observabile și măsurabile forțele și energiile abstracte: „După ce am luat tot în calcul, ceea ce numim realitate apare din adresarea întrebărilor cu răspunsuri de tip da/nu și din consemnarea răspunsurilor evocate de echipament; pe scurt, [...] tot ce este fizic are origini informațional-teoretice”[31].

Modalitățile senzoriale ale sistemului nervos generează „răspunsuri evocate de echipament” la întrebările noastre

despre univers. Răspunsurile senzoriale dau naștere fenomenelor – „lucrurilor” – pe care le trăim. Așadar, experiența însăși apare din comunicarea dintre un receptor al informației (un observator participant) și o sursă a informației, dintre ceva/cineva care pune întrebări și răspunsurile pe care le consemnează. Dar tot nu am răspuns la o întrebare: *de unde vine acel ceva/cineva care pune întrebările?*

Chestiunile astea sunt cam complicate. Înainte să continuăm, propun să facem o recapitulare. Am expus trei idei importante. Prima este că informația medie a unui sistem este entropia sistemului respectiv (adică entropia dintr-un sistem indică cantitatea de informație necesară pentru a-i descrie starea fizică). A doua este că sistemele vii trebuie să se opună entropiei. Aceste două fapte înseamnă că *trebuie să minimizăm informația pe care o procesăm*. (Aici mă refer la informație în sensul întâlnit la Shannon, desigur; cu alte cuvinte, trebuie să ne minimizăm *incertitudinea*.) Tot ce voi mai spune în acest capitol și în următoarele două pleacă de la această concluzie simplă, dar uimitoare.

Ajungem astfel la a treia idee importantă pe care am aflat-o până acum: noi, sistemele vii, ne opunem entropiei prin mecanismul homeostazei. Pe scurt, primim informație despre supraviețuirea noastră probabilă *punând întrebări* (adică măsurând) despre starea noastră biologică în contextul unor evenimente aflate în desfășurare. Cu cât răspunsurile sunt mai incerte (deci cu cât conțin mai multă informație), cu atât mai rău pentru noi – asta înseamnă că nu ne îndeplinim obligația homeostatică de a ocupa stări limitate (stările noastre preconizate).

Natura întrebărilor pe care le punem este determinată

parțial de *specia* căreia îi aparținem. Rechinii pot să respire în apă, oamenii nu. Așadar, avem nevoi diferite și ne așteptăm să ocupăm stări diferite. Asemenea cerințe sunt determinate prin selecție naturală. Homeostaza înseamnă să rămâi în nișa în care te-a adus evoluția speciilor. De aceea, fiecare specie trebuie să-și pună întrebări precum „Pot să respir aici?”. Supraviețuirea noastră depinde de răspunsurile pe care le primim.

Apropo, de ce ar trebui să ne gândim la necesităţile biologice ca la niște *așteptări*? Această exprimare poate părea ciudată în acest punct, dar indică o continuitate profundă despre care vom vedea mai jos că e importantă. Dacă te ajută, încearcă să privești din perspectiva evoluției înseși, și nu a unei creaturi individuale. Selecția naturală a înzestrat fiecare specie pentru nișa sa ecologică: supraviețuirea fiecărei creaturi depinde numai de lucrurile care se găsesc de fapt fără probleme în habitatul său natural. Deci avem nevoie de aer *deoarece* ne putem aștepta să existe.

Acum pot să revin la întrebarea profundă formulată mai sus: de unde vin observatorii participanți? Cu alte cuvinte: cum și de ce, în termeni fizici, apare adresarea unor întrebări?[32]

Iată un foarte scurt istoric al ideii de autoorganizare. Primul care a utilizat acest termen este Immanuel Kant, în opera sa din 1790 *Critica puterii de judecare*. Kant afirmă că ființele au „scopuri” și „obiective” intrinsece care pot fi adevărate numai dacă mecanismele lor componente sunt simultan atât scopuri, cât și mijloace. Asemenea entități „teleologice” (adică entități cu scopuri și obiective intrinsece), spune Kant, trebuie să aibă comportament intenționat: „Numai în aceste condiții și în acești

termeni poate un asemenea produs să fie o ființă organizată și autoorganizată și, ca atare, poate fi numită scop fizic”. Kant considera că modul în care asemenea ființe ar putea să apară e dincolo de puterea explicativă a științei: nu putea niciodată să existe „un Newton al firului de iarbă”.

Apoi Darwin a descoperit selecția naturală. După cum știm acum, selecția naturală dă naștere obiectivelor și scopurilor intrinsece ale *supraviețuirii* și *reproducerii*. Acestea se dovedesc a fi amândouă manifestări ale autoorganizării[33]. Odată cu ideea lui Darwin, chestiunea originii și alcătuirii ființelor teleologice a devenit încadrabilă în domeniile științei[34]. Nu a mai rămas decât să li se deslușească detaliile.

Un alt pas important a fost făcut la mijlocul secolului XX, când Norbert Wiener, matematicianul care a fost întemeietorul „ciberneticii”, a adăugat ideea de *feedback* la perspectiva lui Shannon asupra informației. Conform lui Wiener, un sistem își poate atinge obiectivul („starea de referință”) dacă primește feedback despre consecințele acțiunilor sale. Feedbackul include semnalele de eroare – ce măsoară *deviațiile* de la starea de referință – care ar fi folosite pentru ajustarea acțiunilor sistemului, astfel încât să-și mențină cursul. Astfel, homeostaza se dovedește a fi un caz specific al unui principiu cibernetic mai general: e un fel de feedback negativ.

William Ross Ashby a folosit această idee de feedback, în combinație cu fizica statistică prezentată mai sus, ca să arate modul în care autoorganizarea se desfășoară natural[35]. Ashby a arătat că multe sisteme dinamice complexe evoluează automat spre un *punct de stabilizare*, pe care l-a descris ca pe un „punct de atracție” într-un „bazin” de stări învecinate. Evoluția ulterioară a unor asemenea sisteme tinde apoi *să ocupe stări*

limitate.

Sper că această tendință de a ocupa stări limitate îți sună familiar: nu este nimic altceva decât o tendință de a te opune entropiei. Conform lui Friston, tocmai această tendință declanșează forme tot mai elaborate de autoorganizare. După ce a creat această posibilitate între subsistemele din supa primordială pe care a simulat-o, descrisă la începutul acestui capitol, le-a observat comportamentul desfăşurându-se în trei etape:

1) în cazul unor parametri pe distanțe scurte, punctele doar au țâșnit în toate direcțiile;

2) în cazul altor parametri, punctele s-au adunat în structuri stabile asemănătoare unor cristale;

3) iar în cazul altora, au manifestat comportamente mai complexe: după ce s-au adunat, s-au înghesuit neobosite unele în altele, ocupând poziții specifice în cadrul unei structuri dinamice.

Iată descrierea făcută de Friston (nu-ți bate capul cu limbajul tehnic, vreau doar să transmit o impresie vizuală a ceea ce a văzut):

> Aceste comportamente variază de la comportamente ca de gaz (unde subsistemele uneori se apropie suficient încât să interacționeze) la un focar de activitate, când subsistemele sunt nevoite să se întâlnească în partea de jos a sursei potențiale. În acest regim, subsistemele se apropie suficient de mult încât legea inversului pătratului să le distrugă – amintind de coliziunile particulelor subatomice în fizica nucleară. La anumite valori ale parametrilor, aceste

evenimente sporadice și critice pot să facă dinamica nonergodică, cu fluctuații impredictibile de mare amplitudine care nu se stabilizează. În alte regimuri apare o structură mai cristalină, cu interacțiuni mute și entropie structurală (configurativă) scăzută. Totuși, în cazul celor mai multe valori ale parametrilor, comportamentul ergodic apare pe măsură ce ansamblul se apropie de punctul său de atracție global aleatoriu (de obicei, după aproximativ 1.000 s): în general, inițial subsistemele se resping reciproc (în mare parte ca în ilustrările Big Bangului), apoi cad înapoi spre centru, găsindu-se unul pe altul în timp ce se adună. Apoi, interacțiunile locale mediază o reorganizare, în care subsistemele sunt trecute dintr-o parte în alta (uneori spre periferie) până când vecinii se îngrămădesc ușor unul în altul. În ce privește dinamica, sincronizarea temporară poate fi văzută ca niște valuri de izbucnire dinamică [...]. Pe scurt, mișcarea și dinamica electrochimică seamănă foarte mult cu o supă în clocot (nu cu mult deosebită de erupțiile solare de pe suprafața soarelui) – dar are ea oare mai multă autoorganizare de-atât?[36]

Răspunsul la ultima lui întrebare s-a dovedit a fi pozitiv. Apare o structură dinamică complexă în care subsistemele dense, după ce s-au separat de mediul lor înconjurător, formează *structuri stratificate concentric*, fiecare având un centru interior și o suprafață exterioară și fiind împărțită suplimentar în două substraturi (vezi figura 13). Substraturile suprafeței împărțite manifestă tipare foarte interesante de interacțiune cu centrul interior și, respectiv, cu mediul înconjurător. Stările substratului exterior sunt influențate de cele ale mediului exterior și ele, la rândul lor, le influențează pe

cele ale subsistemelor interne, dar această influență nu este reciprocă (altfel spus, componentele interne ale nucleului nu au niciun impact asupra substratului exterior). De asemenea, stările substratului interior sunt afectate cauzal de cele ale nucleului interior, iar ele, la rândul lor, le influențează pe cele ale mediului exterior, dar, din nou, influența nu este reciprocă. Acest aranjament al dependențelor cauzale definește proprietățile a ceea ce se numeşte *zonă Markov*[37].

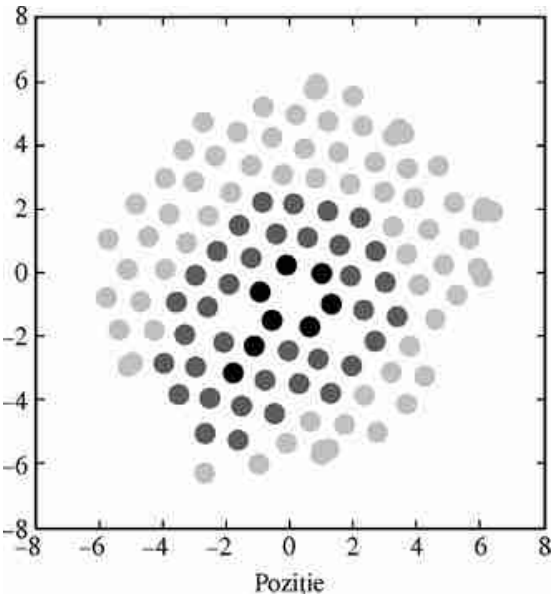


Figura 13. Un sistem cu autoorganizare, cu zona Markov aferentă. În această imagine, nucleul interior al sistemului este ilustrat prin punctele negre, iar straturile sale înconjurătoare prin cele de culoare gri-închis: zona. Punctele gri-deschis sunt exterioare sistemului. (Reprezentarea originală a lui Friston diferența substraturile zonei. De precizat, întrucât culorile lui sunt menționate în text: punctele exterioare erau albastru-deschis, cele din interior erau albastru-închis, cele senzoriale erau violet, iar cele active erau roșii.)

„Zona Markov” este un concept statistic ce separă două seturi de stări. Asemenea formațiuni induc o *împărţire* a stărilor în stări interne și externe, adică într-un sistem și un nu-sistem, astfel încât stările interne sunt izolate de cele externe sistemului. Cu alte cuvinte, stările externe pot fi „simțite” numai *indirect* de cele interne ca stări ale zonei. Mai mult, o zonă Markov este ea însăși împărţită în subseturi care sunt și în subseturi care nu sunt dependente cauzal (direct) de stările setului exterior. Aceste stări ale zonei sunt numite stări „senzoriale” și, respectiv, „active”.

Astfel, formarea unei zone Markov împarte stările unui sistem în patru tipuri – interne, active, senzoriale și externe –, unde stările externe nu fac parte din entitatea cu autoorganizare. Un aspect crucial, dependențele dintre aceste patru tipuri de stări creează o cauzalitate circulară. Stările externe le influențează pe cele interne prin stările senzoriale ale zonei, iar stările interne se conectează iar la cele externe prin stările lor active. În acest fel, stările interne și cele externe se cauzează reciproc într-un mod circular. Altfel spus, stările senzoriale *furnizează feedback* despre consecințele efectului asupra stărilor externe ale stărilor active, ajustând astfel acțiunile ulterioare ale sistemului.

Dacă asta seamănă cu ciclul percepție/acțiune la organismele vii – și îți amintește de ce am spus în capitolul anterior despre cauzalitatea circulară dintre sistemul reticulat activator, prozencefal și triunghiul de decizie din mezencefal – nu e o întâmplare. Exact asta este valoarea unor astfel de modele abstracte. Relevă formalisme profund regulate care pot fi apoi recunoscute în substraturi de multe tipuri – și ne permit să înțelegem din perspective noi structura organismelor vii.

Odată ce începi să cauți, vezi zone Markov peste tot. O membrană celulară are proprietățile unei zone Markov, la fel și pielea și sistemul musculoscheletal al corpului în ansamblu – care este el însuși alcătuit din celule. La fel și fiecare organit, organ și sistem fiziologic. Așadar, creierul (de fapt, întregul sistem nervos) – ce reglează celelalte sisteme ale corpului – are o zonă Markov. De fapt, este o *metazonă,* deoarece le înconjoară pe toate celelalte. Sistemele cu autoorganizare pot fi întotdeauna alcătuite din sisteme cu autoorganizare mai mici – nu până la ultimul element, dar cu siguranță până la unele chinuitor de mici. Aceasta este materia primară a vieții: miliarde de homeostate mici înfăşurate în zonele lor Markov.

Ne apropiem de un răspuns la întrebarea pe care am pus-o mai sus: de ce și cum, în termeni fizici, apare formularea unor întrebări? Încă nu am ajuns acolo, dar, din ce am spus până acum, putem să concluzionăm că însăși ipseitatea unui sistem dinamic complex este constituită de zona sa. Asemenea sisteme cu autoorganizare apar separându-se de orice altceva. Apoi, nu pot decât să își consemneze propriile stări: lumea nu-sistem nu poate fi „cunoscută” decât indirect, prin intermediul stărilor senzoriale ale păturii sistemului. Eu consider că aceste proprietăți ale autoorganizării sunt de fapt precondițiile esențiale pentru subiectivitate.

Se impune o distincție. Tocmai caracteristica de autoconservare a unor astfel de sisteme, tendința de a se separa de mediile lor și apoi de a-și menține activ existența, constituie baza elementară a *ipseității*. Și tocmai natura izolată a unor astfel de sisteme, faptul că pot să consemneze numai lumea nu-individuală prin intermediul stărilor senzoriale ale propriilor zone, constituie baza elementară a *subiectivității* – „punctul de

vedere” al unui sine sechestrat.

Desigur, asta nu înseamnă că fiecare sistem cu autoorganizare are o subiectivitate *capabilă de a simţi*. Încă suntem departe de a putea identifica proprietățile specifice pe care un sistem cu autoorganizare trebuie să le manifeste înainte de a putea deveni conștient. Totuși, chiar dacă nu putem vorbi de conștiință încă, se pare că am dat peste un prototip fizic pentru problema altor minți. Însăși natura unei zone Markov constă în a determina o împărţire a stărilor în „sistem” și „nu-sistem”, astfel încât stările nu-sistem sunt *ascunse* de interiorul sistemului și invers[38].

Să revenim la supa primordială a lui Friston, unde lucrurile devin și mai ciudate. După ce a generat în mod spontan un sistem cu autoorganizare dinamic și complex, Friston a testat dacă acest ansamblu permite *predicții* cu privire la stările externe pornind de la stările interne ale sistemului. Dacă permite, a afirmat Friston, asta ar sugera că stările interne ale unui sistem își *modelează*, în timp, stările externe și s-ar putea spune și că *reprezintă* în interiorul lor acele evenimente externe. Știu că pare magie, dar asta înseamnă pur și simplu că sistemul se adaptează la tiparele evenimentelor externe. (Ca să simplific mult ideea: așa se explică faptul că poți prezice direcția obișnuită a vântului dintr-o zonă judecând după înclinația copacilor de acolo atunci când nu bate vântul. Înclinația copacilor „reprezintă” direcția obișnuită a vântului, deoarece copacii au crescut sub acel unghi ca să se adapteze la vânt.)

Friston a examinat statutul funcțional al subsistemelor interne ale organismului pe care l-a simulat și a descoperit tocmai această capacitate predictivă:

> Dinamica internă care prezice [un eveniment extern] pare să apară în fluctuațiile sale înainte de evenimentul în sine – după cum s-ar anticipa dacă evenimentele interne modelează evenimentele externe. E interesant că subsistemul cel mai bine prezis a fost cel aflat la cea mai mare depărtare de stările interne. Acest exemplu ilustrează că stările interne deduc sau consemnează evenimente îndepărtate într-un mod ce nu e cu totul diferit de perceperea unor evenimente auditive prin undele sonore – sau modul în care peștii simt mișcarea în mediul lor. Subsistemele a căror mișcare a putut fi prezisă cu un grad mare de precizie sunt cele mai semnificative de la periferia ansamblului, unde ansamblul are cea mai mare libertate de mișcare. Aceste mișcări sunt cuplate la stările interne – prin zona Markov – prin sincronism generalizat[39].

După ce a observat acest sincronism, prin care stările interne ale sistemului au modelat evenimente îndepărtate sub aspect fizic, Friston a concluzionat că stările interne manifestă „inferență”.

Aceasta se dovedește a fi cea mai semnificativă proprietate a unor asemenea sisteme. Zona Markov conferă stărilor interne ale sistemelor cu autoorganizare o capacitate de a reprezenta sub aspect probabilistic stări externe ascunse, astfel încât sistemul poate *să infereze cauzele ascunse ale propriilor stări senzoriale*, ceea ce seamănă cumva cu funcția percepției. La rândul său, această capacitate permite sistemului *să acționeze în mod intenționat* asupra mediului exterior, în baza stărilor sale interne – acțiuni ce sunt similare activității motoare.

În acest fel, sistemul se menține și se reînnoiește în fața perturbațiilor externe[40]. Simplul fapt că este un sistem cu autoorganizare este suficient pentru a le conferi un scop lui și fiecăreia dintre părțile sale, iar aceasta este funcția stărilor active ale zonei: ele manipulează mediul pentru a menține integritatea sistemului. Ceea ce înseamnă că, alături de un sine închis, un punct de vedere subiectiv, un obiectiv și capacitatea și de a simți, și de a acționa, simplul fapt că există o zonă Markov aduce cu sine ceva similar unei *intenționalități*[41]. Poate că nu pare deosebit de impresionant în forma în care se vede în simulările lui Friston – punctele albastru-deschis (exterioare) le influențează pe cele albastru-închis (interioare) prin cele violet (senzoriale), iar punctele albastru-închis se conectează iar la cele albastru-deschis prin cele roșii (active). Cu toate acestea, sper că înțelegi de ce sistemele biologice cu autoorganizare trebuie să infereze cauzele ascunse ale stărilor lor senzoriale – chiar dacă deocamdată în mod nonconștient. Dacă nu ar face-o, ar înceta să existe. Ele nu au cum să nu modeleze dependențe cauzale din lume, astfel încât acțiunile lor din lumea respectivă să le asigure supraviețuirea.

De aici vine conceptul de „stări preconizate” și din acest motiv sistemele biologice cu autoorganizare sunt homeostatice. Homeostaza pare să fi apărut odată cu autoorganizarea. Stările senzoriale și active ale unei zone Markov nu sunt decât receptori și efectori ai unui sistem cu autoorganizare, iar modelul stărilor externe pe care îl generează este centrul său de control.

Sistemele biologice cu autoorganizare *trebuie* să-și testeze modelele lumii și, dacă lumea nu oferă răspunsurile pe care le așteaptă, ele trebuie să facă urgent ceva diferit, ca să nu moară.

Prin urmare, deviațiile de la stările preconizate sunt o formă elementară a „răspunsurilor evocate de echipament” ale lui Wheeler. Așa apare formularea întrebărilor; *autoorganizarea îi face pe observatorii participanți să existe*. Întrebarea pe care și-o pune întotdeauna un sistem cu autoorganizare este, pur și simplu: „Supraviețuiesc dacă fac asta?”. Cu cât răspunsul e mai nesigur, cu atât mai rău pentru sistem.

Relația dintre stările active ale unei zone Markov și autoconservare, prin homeostază, este ilustrată cel mai bine dacă arătăm ce s-a întâmplat când Friston *a cauzat leziuni* zonei sistemelor cu autoorganizare în simulările lui experimentale. A făcut acest lucru împiedicând selectiv stările senzoriale ale zonei să-i influențeze stările active (vezi figura 14, graficele b, c și d). În lipsa stărilor active obișnuite ale zonei Markov, a apărut haos entropic, iar sistemul s-a disipat rapid. Adică a încetat să existe.

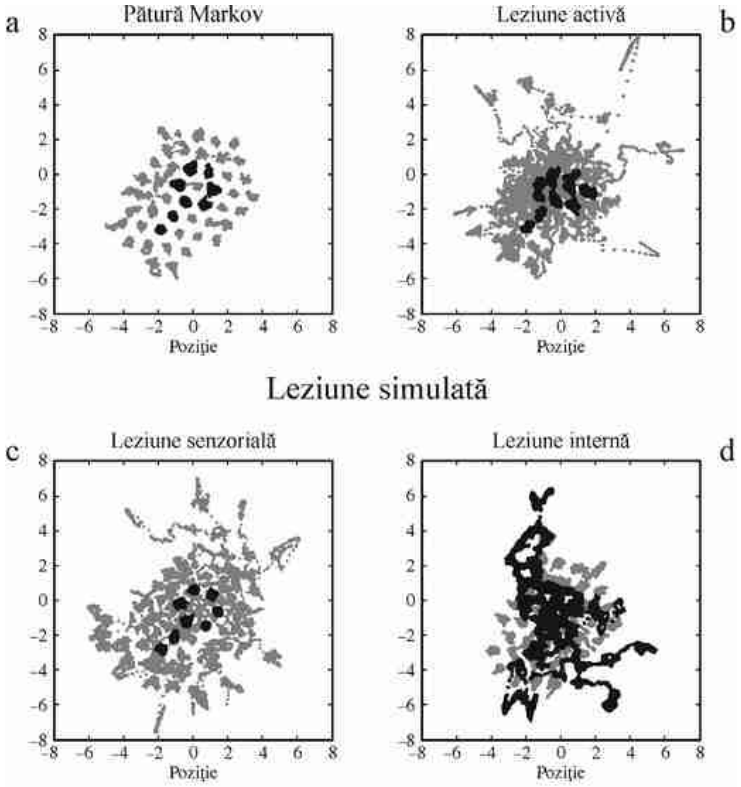


Figura 14. Efecte entropice ale unor leziuni ușoare la nivelul zonei Markov a unui sistem cu autoorganizare

Ceea ce tocmai am ilustrat în termeni formali (folosind concepte din fizica statistică) este structura de bază a oricărui sistem cu autoorganizare dinamic. Friston rezumă relațiile legitime după cum urmează: „*Orice sistem dinamic ergodic aleatoriu ce are o zonă Markov va părea că își menține în mod activ integritatea structurală și dinamică*”[42]. (Un sistem „ergodic” ocupă stări limitate.) Acest tip de activitate îndeplinește criteriile lui Kant menționate mai sus: o zonă Markov este atât scopul, cât și mijlocul prin care un sistem cu autoorganizare persistă în timp – iar acest lucru se întâmplă în mod natural. Astfel, asemenea sisteme par să aibă o minte proprie, chiar dacă foarte primitivă (și nonconștientă).

Dacă această afirmație este corectă, atunci ar trebui să fie posibil să identificăm apariția autoorganizării în orice ansamblu arbitrar de subsisteme cuplate cu interacțiuni pe distanță mică. Exact asta a arătat experimentul cu supa primordială al lui Friston. Din acest experiment, el a extras patru proprietăți fundamentale ale *tuturor* sistemelor biologice cu autoorganizare:
1) sunt ergodice;
2) au o zonă Markov;
3) manifestă inferență activă;
4) se autoconservă.

Ce anume este principiul energiei libere ce dă titlul acestui capitol? În primul rând, se știe că este dificil de explicat. Iar asta în mare parte din cauza ecuațiilor lui Friston, care, după cum am descoperit când am reușit să înțeleg „Life as we know it", sunt şi greu de priceput, și opace. Așa că voi încerca să explic principiul numai cu ajutorul cuvintelor. Ecuațiile pot fi întotdeauna exprimate în cuvinte, deoarece nu sunt decât afirmații despre niște relații. Acestea fiind spuse, întrucât ecuația de bază îți explică scopul şi obiectivele tale principale în viață, precum și pe cele a orice altceva care a trăit vreodată, ar trebui probabil să o vezi în forma ei consacrată măcar o dată. Iat-o:

$$A = U - TS$$

unde A înseamnă energia liberă, U este energia internă totală, T temperatura, iar S entropia.

Ce înseamnă asta? În termodinamică, energia liberă a unui

sistem este egală cu cantitatea totală de energie conținută în sistem minus partea folosită deja într-o activitate efectivă, deci care nu este liberă[43]. Așadar, ecuația spune doar: „Energia liberă este egală cu energia internă totală minus energia utilizată deja”. Ce ar putea fi mai simplu? Energia liberă este ceea ce rămâne după ce îndepărtezi energia care nu este liberă (adică atunci când îndepărtezi „energia latentă”)[44].

Ecuația pe care tocmai am descris-o cuantifică cu acuratețe energia liberă în contexte termodinamice elementare, dar nu şi în contexte chimice, unde trebuie admisă o toleranță pentru moleculele suplimentare ce sunt formate de unele procese la diverse temperaturi și presiuni. De aceea, chimiștii folosesc o versiune ușor diferită a ecuației[45]. Pentru a le deosebi, tipul termodinamic clasic de energie liberă este numită „energie liberă Helmholtz” (de la Hermann von Helmholtz, un membru important al Societății de Fizică din Berlin), iar tipul ansamblului chimic este numită „energie liberă Gibbs”.

Friston folosește o a treia versiune a aceleiași ecuații, pentru a cuantifica energia liberă în contexte *informaționale*. Voi numi „energie liberă Friston” acest tip de energie liberă. Ecuația relevantă spune că „energia liberă Friston este egală cu energia medie minus entropia”[46]. Aici, „energie medie” înseamnă probabilitatea *preconizată* ca un eveniment să se întâmple conform unui model, iar „entropie” înseamnă incidența *reală* a concretizării modelului. Așadar, energia liberă Friston este diferența dintre cantitatea de informații pe care te aștepți s-o obții dintr-un eșantion de date – dintr-o secvență de evenimente – și cantitatea de informații pe care chiar le obții. (Să ne amintim că entropia unui sistem predictiv este informația sa medie, iar creșterea informației înseamnă

descreșterea probabilității[47].) Prin urmare, ecuația „energia liberă Friston este egală cu energia medie minus entropia” spune, practic, același lucru ca ecuația „energia liberă Helmholz este egală cu energia internă totală minus energia care nu este disponibilă să acționeze”[48]. Asta se întâmplă deoarece energia liberă Friston este analoagă cu energia liberă Helmholtz, unde există un schimb de *informații*, spre deosebire de un schimb *termodinamic* între sistem și mediul său[49].

Dacă sistemele biologice trebuie să-și minimizeze entropia, iar entropia este informație medie, atunci ele trebuie să mențină fluxul de informații pe care le prelucrează la un nivel minim. Trebuie să minimizeze evenimentele nepreconizate. Acest lucru este cunoscut sub denumirea tehnică de „surprindere”. La fel ca entropia, surprinderea este o funcție descendentă a probabilității: pe măsură ce probabilitatea scade, surprinderea crește[50]. Surprinderea măsoară cât de improbabil a fost un eveniment; entropia măsoară cât de improbabil se preconizează că va fi (în medie)[51]. Așadar, surprinderea, ca şi entropia din model, este un lucru rău pentru organismele vii. Nu vreau să ne pierdem în detalii tehnice[52]. Vreau să spun doar că, la nivelul cel mai simplu, ești într-o stare de surprindere dacă ieși în afara setului de stări în care *te aştepţi* să te afli din punct de vedere biologic (de exemplu, sub sau peste 36,5-37,5ºC sau respirând sub apă), tocmai pentru că există o probabilitate mică să fii în acea stare.

Sistemele cu autoorganizare trebuie să minimizeze fluxul informațional, deoarece creșterea cererii de informații presupune creșterea incertitudinii în modelul predictiv. Incertitudinea aduce surprize, care sunt rele pentru noi, ca sisteme biologice, întrucât pot fi periculoase. Dar cum putem să

minimizăm surprizele prin minimizarea fluxului informațional? Oare asta nu înseamnă să ne prefacem că realitatea nu există? Răspunsul este nu. Energia liberă Friston este un atribut cuantificabil al diferenței dintre *modul în care lumea este modelată* de un sistem și *modul în care lumea se comportă de fapt*. Așadar, trebuie să minimizăm această diferență. Modelul de lume al unui sistem trebuie să se potrivească cât mai mult cu lumea reală, ceea ce înseamnă că trebuie să minimizeze diferența dintre datele senzoriale pe care le prelevează din lume și datele senzoriale ce au fost prognozate de modelul său. Asta va maximiza „informarea reciprocă" dintre lume și model, ceea ce minimizează incertitudinea.

O cale de a face acest lucru constă în *îmbunătățirea modelului de lume al sistemului*. Erorile de predicție pot fi transmise înapoi modelului generativ, astfel încât acesta să genereze predicții mai bune data următoare. Mai puține erori făcute înseamnă mai puține erori de transmis modelului, ceea ce înseamnă mai puțin flux informațional. Prin urmare, „informarea reciprocă" este un produs al comunicării, al formulării și răspunderii la întrebări: se comportă lumea cum am prevăzut, da sau nu?

Dar, întrucât sistemele biologice ca mine și ca tine sunt izolate de lume prin zona Markov a fiecăruia, nu ne putem compara direct modelele cu felul în care este lumea cu adevărat. De aceea, trebuie să ne aducem în cap întregul proces de minimizare a surprinderii, devenind atât „sursa", cât și „receptorul" informației ce vine din întrebările pe care le-am pus.

Facem acest lucru prin măsurarea entropiilor relative – prin

cuantificarea decalajului dintre stările senzoriale prezise de o acțiune și stările senzoriale ce apar efectiv din acțiunea respectivă. De aici apare cantitatea numită energie liberă Friston, care este întotdeauna o valoare pozitivă mai mare decât surprinderea propriu-zisă.

Dovezile senzoriale (primite sub forma aşa-numitelor *spike-trains* generate în capul nostru – miliarde de unu și de zero) sunt singurele date pe care le putem obține. Din aceste date trebuie *să inferăm* structura cauzală a lumii. Pentru că suntem ființe cu o zonă Markov, suntem obligați să ne bazăm pe chestiuni precum distribuțiile de probabilitate, și nu pe adevăruri absolute. Iată motivul pentru care este util să știm că energia liberă Friston este întotdeauna mai mare decât surprinderea – permite creierului uman să aproximeze adevăruri necognoscibile folosind calcule statistice.

După cum am văzut în experimentul lui Friston cu supa primordială, modelele generative apar cu sisteme cu autoorganizare. Din acest motiv, ele sunt numite uneori sisteme „care se autovalidează”, deoarece modelează lumea în funcție de viabilitatea lor, apoi caută dovezi pentru modelele lor. E ca și cum ar spune nu „Gândesc, deci exist”, ci „Exist, deci modelul meu este viabil”[53]. Modelul propriu fiecărui sistem biologic este determinat parțial de *specia* sa, cum am explicat deja. Tu – ca ființă umană – te aștepți să ajungi în stări foarte diferite de cele ocupate în mod normal de un rechin. E foarte improbabil să ajungi să respiri apă la sute de metri adâncime în ocean, dar deloc improbabil pentru un rechin. Deci un anumit set de stări senzoriale este mai mult sau mai puțin surprinzător – mai mult sau mai puțin improbabil – în funcție de specia organismului care se află în ele.

Testul unui model bun al sinelui-în-lume constă în a vedea în ce măsură permite sistemului autogenerat să interacționeze cu lumea în moduri care îl mențin în limitele sale viabile. Cu cât aceste interacțiuni sunt mai bune, cu atât mai mică va fi energia sa liberă. Cu cât este mai mică energia sa liberă, cu atât mai multă din energia sistemului este transformată în lucru mecanic util, pentru autoconservare[54]. Astfel, principiul energiei libere explică în termeni matematici modul în care sistemele vii ca mine și ca tine se opun celei de-a doua legi a termodinamicii printr-un lucru mecanic care menține homeostaza.

De asemenea, explică alt mod în care sistemele cu autoorganizare se autovalidează (aproape că aș vrea să spun că sunt introspective): ele sunt obligate să-și pună întrebări *lor înseși* despre propriile stări. Mai exact, sunt obligate să întrebe: „Ce se întâmplă cu energia mea liberă dacă fac asta?”. Răspunsul la această întrebare va determina întotdeauna ce urmează să facă sistemul în continuare, într-un interval adecvat[55]. Acesta este mecanismul cauzal din spatele tuturor comportamentelor voluntare[56].

În mod surprinzător, nu am ajuns încă la capătul puterilor psihice remarcabile ce se manifestă în supa primordială a lui Friston. Am văzut că până și cele mai simple sisteme cu autoorganizare au individualități nonconștiente și subiectivitate, intenționalitate și un scop (şi anume supraviețuirea). Ele percep și, dacă este necesar, acționează. La această listă impresionantă putem adăuga încă un trofeu. Se dovedește că ele manifestă un fel de raționalitate. Putem să spunem chiar că sunt bayesiene.

Thomas Bayes a fost un preot și teolog englez al cărui nume încă nu a fost uitat exclusiv datorită ideilor sale despre probabilitate, pe care nu s-a ostenit niciodată să le publice. Bayes ne-a învățat (într-o lucrare publicată postum, în 1763) că trebuie să folosim *dovezi curente* împreună cu *cunoștințe de fond* ca să facem și să revizuim *cele mai bune conjecturi* despre lume. Cu alte cuvinte, folosind terminologia de-acum familiară a lui Friston, trebuie să prelevăm mostre senzoriale, să le comparăm cu predicțiile derivate din modelele noastre generative și să ne actualizăm credințele în mod corespunzător.

Iată formula standard[57]:

$$P(A|B) = \frac{P(A|B)P(A)}{P(B)}$$

Exprimată în cuvinte, această teoremă spune: „Raportul probabilităților pentru două ipoteze condiționate de un număr de date este egal cu raportul probabilităților lor condiționale înmulțit cu gradul în care prima ipoteză o depășește pe a doua ca predictor al datelor”.

Mai simplu: dată fiind o ipoteză urmată de niște dovezi, trebuie să revizuiești probabilitatea ipotezei analizându-i *verosimilitatea* coroborată cu *probabilitatea* sa *anterioară*. „Verosimilitatea” ipotezei este gradul de potrivire dintre ce prezice și dovezile obținute efectiv, iar „probabilitatea anterioară” constă în cunoașterea ta de fond despre ipoteză (adică probabilitatea ei chiar înainte de a lua în considerare dovezile noi). Rezultatul este *probabilitatea posterioară* a ipotezei. Atunci, date fiind două convingeri concurente, conjectura ta cea mai bună este cea care are cea mai mare probabilitate posterioară. (Vezi figura 16 de la p. 207.)

De exemplu: ești la aeroportul din Cape Town și vezi pe cineva venind dintr-o cursă de la Johannesburg care seamănă

cu prietena ta Teresa. Ipoteza ta e că aceasta este Teresa. „Verosimilitatea” ei este probabilitatea că vezi pe cineva care arată așa, dat fiind că ea este Teresa. Apoi îți amintești că Teresa locuiește la Londra, ceea ce reduce „probabilitatea anterioară” ca aceasta să fi ea. Concluzia ta („probabilitatea posterioară”) este că te uiți la altcineva care doar seamănă cu Teresa[58].

Cel mai important aspect al teoremei lui Bayes pentru scopurile neuroștiințelor este că explică modul în care funcționează în realitate inferența perceptivă – un proces inconștient – și modul în care funcționează în realitate transmiterea semnalelor în procesarea senzorimotoare. Circuitele cerebrale calculează *concret* distribuțiile probabilității anterioare, apoi trimit mesajele predictive unor neuroni senzoriali, într-un efort continuu de atenuare a semnalelor de intrare; iar percepția implică *concret* comparații între distribuțiile prezise și cele reale, rezultând astfel calcule ale probabilității posterioare. Inferențele rezultate constituie ceea ce *este*, *concret*, percepția[59]. Percepția este un efort de a autogenera semnalele senzoriale de intrare și, prin aceasta, de a le suprima. Acesta este motivul pentru care atât de mulți specialiști în neuroștiințe vorbesc astăzi de „creierul bayesian”.

Să ne amintim ce am afirmat mai sus despre relația dintre „informații” și lumea materială: deși informațiile nu pot fi văzute și atinse concret, nu există niciun dubiu în privința existenței lor. Comportamentul sistemelor fizice este *determinat* de fluxurile informaționale. De aici, minimizarea energiei libere Friston minimizează în același timp energia liberă Gibbs și energia liberă Helmholtz, iar asta pentru că minimizarea erorii de predicție minimizează fluxul informațional, iar

reducerea fluxului informațional reduce *consumul metabolic* din partea creierului și a corpului în ansamblu[60]. Acest lucru nu se întâmplă numai pentru că activitatea cerebrală consumă atât de multă energie (20% din total), ci și pentru că minimizarea energiei libere statistice din creier reglează schimburile energetice fiziologice dintre corp și lume[61]. Aceste schimburi presupun de asemenea consum metabolic. Astfel, se vede că creierul predictiv este „leneș” (pe termen lung): atent la orice ocazie de a obține mai mult cu mai puțin efort[62].

Asta este o explicație minimalistă a ceea ce face creierul. Totuși, ca să mențină organismul viu, creierul trebuie să facă mai mult decât să conserve resursele de energie – trebuie să țină seama și de numeroasele nevoi biologice (pe lângă echilibrul energetic) despre care am discutat în capitolul 5, aproape toate obligându-ne să efectuăm lucru mecanic în lumea exterioară. Vom vedea că diversitatea nevoilor ce ne caracterizează pe noi, ca organisme complexe, nu poate fi separată de conștiință.

Am aflat că suprimarea erorii de predicție este mecanismul esențial al homeostazei. Prin urmare, minimizarea energiei libere devine sarcina de bază a tuturor sistemelor homeostatice. Am văzut că, de fapt, ecuația energiei libere elaborată de Friston este o reformulare, în termeni cuantificabili, a definiției date de Freud „pulsiunii” – „o măsură a cererii de travaliu ce e impusă psihicului ca urmare a legăturii sale cu ceea ce e corporal” –, măsură considerată de Freud necuantificabilă. Acum o putem cuantifica. Forța motrice fundamentală din spatele comportamentului volițional al tuturor formelor de viață este că ele sunt obligate să-și minimizeze energia liberă. Acest principiu guvernează tot ce

fac.

În formularea succintă a lui Friston, principiul energiei libere impune că toate cantitățile care se *pot* schimba – cu alte cuvinte, care fac parte din sistem – se vor schimba *cu siguranță* pentru a minimiza energia liberă[63]. De fapt, această formulare e mai mult decât succintă, este o lege. Propun să o numim „legea lui Friston”: toate cantitățile dintr-un sistem cu autoorganizare ce se pot schimba se vor schimba ca să minimizeze energia liberă. Așadar, iat-o. Înarmați cu această cunoaștere, tot ce numim viață psihică devine încadrabil în matematică.

Suntem aproape pregătiți să aflăm ce este conștiința, în termeni formali și mecaniciști. Dar mai întâi, o fabulă[64].

8

O ierarhie predictivă

Eve Periaqueduct, inginer constructor, este angajată ca să prevină și să repare fisurile dintr-un baraj municipal. Ea nu știe că obiectivul mai profund al jobului ei este să asigure o sursă sigură de apă și energie pentru un sat din apropiere și ca acesta să nu fie luat de viituri. Dar nu are nevoie să știe – unica ei sarcină e să minimizeze fisurile barajului. (Poate că își amintește din timpul facultății că ea minimizează entropia barajului. Dar nu are nevoie să-și amintească asta, sarcina ei este foarte practică.)

Angajatorii ei îi asigură echipamentele necesare și o mică echipă de lucrători. Moștenește și un set de instrucțiuni elaborate de predecesorii ei care indică punctele cele mai slabe ale barajului, spunându-i ce să facă și când. Ea și echipa ei întreține și repară barajul cu sârguință, concentrându-se proactiv asupra punctelor slabe și astupând orice fisuri spontane, pe măsură ce apar. Treptat, de-a lungul anilor, îşi dă seama că și unele dintre fisurile neprevăzute apar după niște tipare regulate. Prin urmare, ea *actualizează* instrucțiunile care i-au fost transmise, devenind mai pricepută în anticiparea (și, astfel, prevenirea) fisurilor. Acest lucru reduce costurile.

Întreprinzătoarea Eve își dă seama că tiparele pe termen lung ale fisurilor pe care le consemnează se corelează cu condițiile climatice. Fără să-şi dea seama, consemnările ei *creează un model* al climei locale (adică vremea și consemnările ei implică „informare reciprocă”). Fără să-și propună, ea generează un model al unui aspect al lumii de dincolo de baraj. Există tipare ale vremii, iar ele corespund tiparelor fisurilor.

Pornind de la această descoperire, Eve angajează personal suplimentar ca să înființeze un departament de meteorologie, pe care îl numește unitatea ei de „detecţie meteorologică”. Acest lucru generează un nivel nou în ierarhia echipei sale, aflat într-un loc diferit, justificat de așteptarea că obținerea unor prognoze meteo mai bune va duce la diminuarea costurilor cu reparațiile pe termen lung.

Acest nivel nou face modelul ei predictiv să fie mai sensibil la contexte preconizate. Nu e nevoie ca angajații departamentului de detecţie meteorologică să știe că treaba lor are legătură cu prevenirea fisurilor; ei se concentrează exclusiv asupra sarcinii de a anticipa schimbările vremii. Eve le furnizează o schemă a condițiilor preconizate derivate din instrucțiunile moștenite iniţial și actualizate de ea. De reținut că aceste instrucțiuni nu vizează tiparele preconizate ale fisurilor, ci condițiile climatice preconizate.

Deoarece Eve nu vrea să piardă mult timp verificându-și mesajele, noului departament i se cere să-i transmită rapoarte de feedback numai dacă apar *devieri* de la condițiile preconizate. Le numește rapoarte de „eroare” și le folosește ca să-și actualizeze în continuare schema condițiilor climatice preconizate, pe care o retrimite apoi la stația meteo, știind că astfel va reduce volumul de muncă al angajaților – și, în cele

din urmă, și al ei.

Toate acestea permit departamentului de meteorologie să se axeze eficient pe sarcina trasată de ea. Ca să-și îndeplinească sarcinile, angajații acestuia instalează o serie de instrumente de măsurare a vremii în locuri diferite, unele aflate la mare distanță de baraj. Barometrele, termometrele, pluviometrele și altele asemenea sunt calibrate de echipă astfel încât să trimită semnale stației meteo numai atunci când parametrii pe care îi măsoară (presiunea atmosferică, temperatura, umiditatea etc.) deviază de la limitele preconizate. Aceste limite sunt stabilite conform condițiilor climaterice anticipate. Şi acest lucru eficientizează costurile, întrucât lucrătorii angajați să citească instrumentele (care creează încă un nivel ierarhic) nu trebuie să ajungă decât la cele care transmit semnale „de eroare” stației meteo. Pentru că ține evidența acestor semnale, stația poate să ajusteze periodic limitele preconizate pentru fiecare instrument, astfel *automatizându-și* și mai mult procedurile (adică reducând frecvența semnalelor care impun trimiterea pe teren a unor angajați care citesc și reglează instrumentele).

Unii dintre algoritmii ce rezultă devin destul de subtili, pe măsură ce echipa descoperă că fluctuațiile parametrilor pe care îi măsoară nu sunt neapărat stabile și regulate, ci variază în funcție de *context*. De exemplu: „Dacă presiunea atmosferică e scăzută, trebuie să crești volumul precipitațiilor preconizate, dar numai iarna”. Chiar și așa, nu e nevoie ca angajaţii care citesc și calibrează instrumentele să aibă informații despre sarcina mai amplă a departamentului de meteorologie. Unica lor sarcină este să citească și să calibreze instrumentele, conform instrucțiunilor actualizate pe care le primesc de la biroul meteo. Și mai puțin relevant pentru jobul lor este faptul

că rapoartele departamentului de meteorologie despre devierile de la condițiile climaterice preconizate îi sunt trimise lui Eve, pentru ca ea să poată prezice cu un grad mai mare de precizie tiparele fisurilor și, prin urmare, să ia măsuri mai eficiente de mentenanță a barajului.

De asemenea, sătenii din apropiere folosesc aceste prognoze meteo în alte scopuri, care nu au nicio legătură cu barajul. Astfel, ei au o impresie eronată despre obiectivul real al departamentului – cred că există ca să-i ajute să-și planifice activitățile sociale în aer liber.

Eve Periaqueduct observă, în timp, că tiparul fisurilor din baraj se corelează nu numai cu condițiile climaterice, ci și cu evenimentele seismice. Prin urmare, înființează o a doua echipă specializată, pe care o numește departamentul de „detecţie a cutremurelor”. Acesta se axează exclusiv pe generarea unor modele ale mișcărilor tectonice și prezicerea lor, așa că acest al doilea departament instalează, calibrează, monitorizează și ajustează continuu propriile echipamente „senzoriale” tehnice. De asemenea, ține evidențe complexe – care, ca și în cazul departamentului de meteorologie și al lui Eve însăși, permit noii echipe să automatizeze anumite aspecte ale activității sale și să se axeze numai pe fluctuațiile imprevizibile pe termen scurt. (Și sătenii folosesc prognozele în beneficiul propriu, deși nu aceasta a fost menirea prognozelor.)

Astfel, se conturează gradual o ierarhie predictivă complexă, cu mai multe departamente, fiecare cu subniveluri proprii, ce măsoară diferiți parametri din lumea de dincolo de baraj. Fiecare nivel al ierarhiei urmează numai instrucțiunile predictive actualizate pe care le primește de la nivelul imediat superior și raportează numai devierile de la stările preconizate

ale parametrilor monitorizați la acel nivel. Din punctul de vedere al lui Eve, rapoartele mixte pe care le primește de la departamentele ei senzoriale se contextualizează reciproc. Din când în când, ea trebuie să hotărască ce raport să *prioritizeze*. Până la urmă, dispune de resurse limitate și nu se poate ocupa de toate evenimentele posibile.

Eve se ocupă în continuare de mentenanța barajului în baza programului ei pe termen lung, de la care deviază numai atunci când nu se potrivește cu prognozele combinate pe care le primește de la departamentele senzoriale. La rândul lor, departamentele îi trimit lui Eve rapoarte de feedback numai atunci când eșantioanele de date pe care le culeg deviază de la propriile predicții stabilite de multă vreme. Și tot așa, până la angajații care citesc și reglează instrumentele de măsură.

De asemenea, întregul proces al transmiterii de mesaje și al actualizării programului ce se tot repetă între nivelurile organizației înființate de Eve respectă regula lui Bayes: folosește dovezile curente (mostre senzoriale) împreună cu cunoașterea de fond (ipoteze anterioare) ca să facă *și să-și revizuiască* cele mai bune conjecturi (ipoteze posterioare) despre lume.

Cu timpul, munca lui Eve devine repetitivă și plictisitoare și ea abia așteaptă pensionarea. Își dă seama că își pune problema: „Înainte să plec, aș vrea să construiesc un baraj nou și mai bun”. Așa că sună la conducerea municipală și întreabă: avem cumva un departament „reproductiv”?

Foarte multe funcții cvasipsihice apar la majoritatea celor mai simple sisteme cu autoorganizare, dar, ca să explicăm cum funcționează un creier propriu-zis, trebuie să vedem cum se pot

potrivi aceste sisteme într-o structură generală reciproc avantajoasă. Datorită studiilor lui Friston, este evident că sistemul nervos implementează o ierarhie predictivă iterativă care funcționează în mare măsură ca şi cea stabilită de Eve Periaqueduct de-a lungul timpului. În cele din urmă, numeroasele funcții complexe ale creierului chiar pot fi reduse la câteva asemenea mecanisme simple. După cum scrie Jakob Hohwy:

> Creierul încearcă oarecum cu disperare, dar cu pricepere să țină sub control efectele pe termen scurt și pe termen lung ale cauzelor ambientale asupra organismului, ca să-și mențină integritatea. Procedând astfel, apare implicit o reprezentare bogată, stratificată a lumii. Aceasta este o imagine frumoasă a minții și a locului nostru în natură, care ne îndeamnă la modestie[1].

De reținut că în științele cognitive cuvântul „implicit” înseamnă *inconștient*.

În centrul modelului creierului despre sine în lume sunt generate predicții specifice speciei despre limitele sale viabile (partea din stânga a figurii 15)[2]. Acestea se concretizează în reflexele autonome ce iau forma: „Dacă fac asta, temperatura mea va fi de aproximativ 37ºC”. La nivelul următor (în dreapta), înconjurând acest centru, creierul generează comportamente instinctive (ce iau forma predicțiilor înnăscute pe care le-am descris în analiza emoțiilor de bază). La următorul nivel, generează comportamente involuntare dobândite (din sistemele sale de memorie nondeclarativă pe termen lung). La următorul, generează comportamente voluntare (din sistemele

sale de memorie declarativă pe termen lung). Iar la final, la nivelul de la granița cu exteriorul, generează acțiunile de tip aici-și-acum, cele mai nesigure care „prezic prezentul” (din sistemele sale de memorie pe termen scurt).

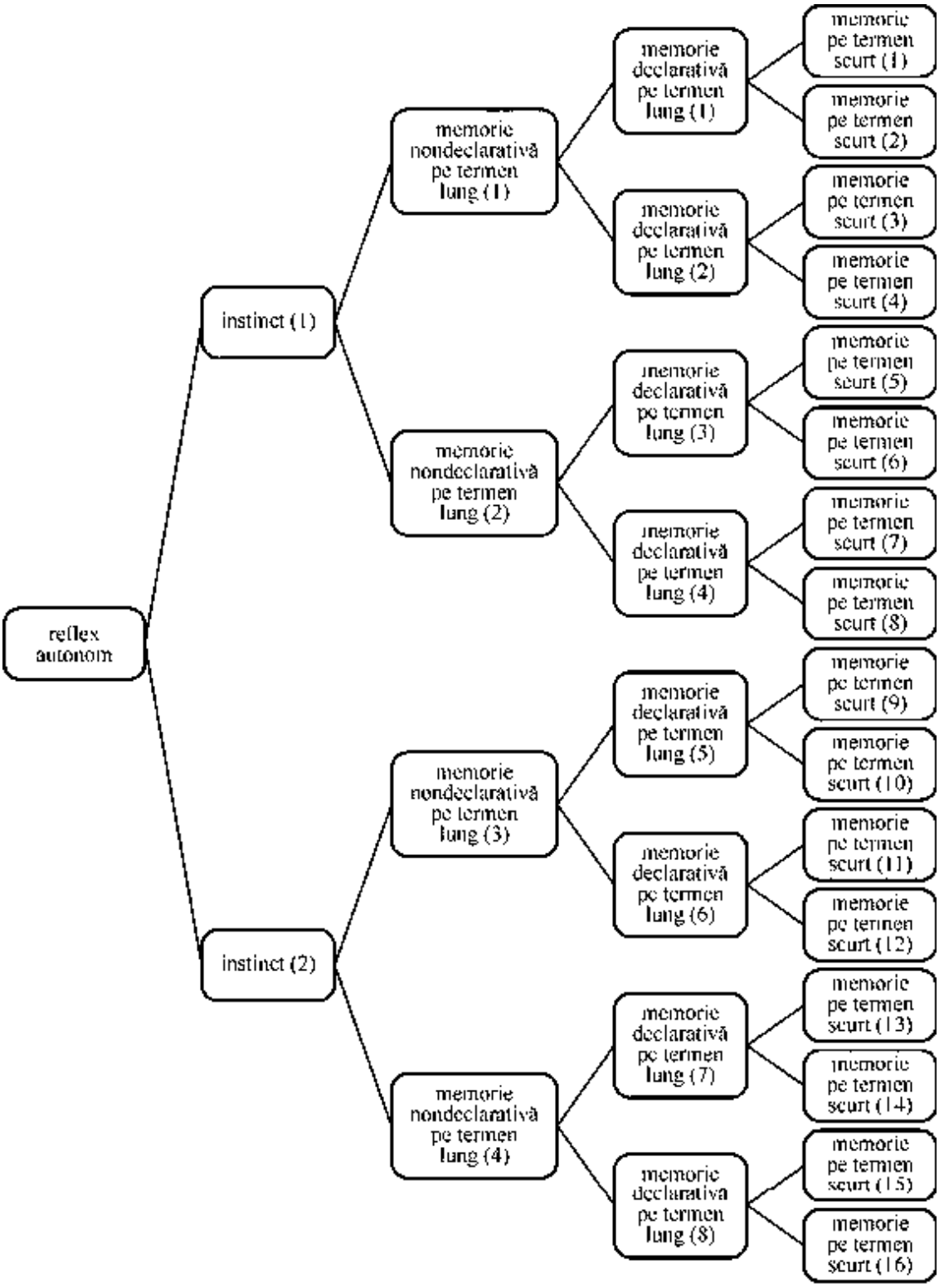

reflex autonom
instinct (1)
instinct (2)
memorie nondeclarativă pe termen lung (1)
memorie nondeclarativă pe termen lung (2)
memorie nondeclarativă pe termen lung (3)
memorie nondeclarativă pe termen lung (4)
memorie declarativă pe termen lung (1)
memorie declarativă pe termen lung (2)
memorie declarativă pe termen lung (3)
memorie declarativă pe termen lung (4)
memorie declarativă pe termen lung (5)
memorie declarativă pe termen lung (6)
memorie declarativă pe termen lung (7)
memorie declarativă pe termen lung (8)
memorie pe termen scurt (1)
memorie pe termen scurt (2)
memorie pe termen scurt (3)
memorie pe termen scurt (4)
memorie pe termen scurt (5)
memorie pe termen scurt (6)
memorie pe termen scurt (7)
memorie pe termen scurt (8)
memorie pe termen scurt (9)
memorie pe termen scurt (10)
memorie pe termen scurt (11)
memorie pe termen scurt (12)
memorie pe termen scurt (13)
memorie pe termen scurt (14)
memorie pe termen scurt (15)
memorie pe termen scurt (16)

Figura 15. O ierarhie predictivă simplificată, extinzându-se de la centrul autonom tonic la periferia senzorimotoare fazică.

Predicțiile circulă dinspre centru spre periferie (de la stânga la dreapta, în această reprezentare); erorile de predicție circulă în direcția opusă.

Simplific lucrurile, desigur: există mult mai multe niveluri în ierarhia predictivă a creierului, nu doar cinci, aranjate în numeroase fluxuri de procesare paralele. Totuși, se întrevăd niște principii generale.

Primul este că creierul conspiră ca să anticipeze și, astfel, „să elucideze" evenimentele din lume. Creierul suprimă semnalele de intrare predictibile, care nu au valoare de informație şi pe care altfel ar trebui să le prelucreze în mod inutil. Pe scurt, fiecare nivel din ierarhia sa primește numai informaţiile nepreconizate, demne de a fi știute, transmise de la nivelul imediat superior. Aceste rapoarte de feedback sunt erori de predicție.

Al doilea principiu general este că această ierarhie se desfășoară la scări temporale și spațiale progresiv mai mici. Predicțiile centrului se aplică în toate situațiile, iar cele mai periferice sunt temporare și focale. O secvență predictivă se desfășoară de la nucleii care monitorizează corpul localizați în trunchiul cerebral și în diencefal[3], prin ganglionii bazali și sistemul limbic, prin neocortex, spre receptorii senzoriali specifici tipurilor de senzații localizați în organele terminale (de exemplu, în bastonașele sau conurile retinei), care au câmpuri receptoare foarte înguste. La periferie, acuratețea pe termen scurt și complexitatea prevalează în detrimentul generalizabilității pe termen lung, de care se bucură predicțiile mai profunde.

Al treilea principiu este conex: există o ierarhie a plasticității exprimată în predicțiile centrului care nu se pot schimba și cele periferice care se pot schimba și se schimbă; ele sunt supuse unei actualizări instantanee, nivelurile intermediare generând un grad intermediar de plasticitate. Cu alte cuvinte, „centrul de control” al homeostatului creierului (modelul propriului sine) se actualizează constant – deși manifestă o rezistență tot mai mare la schimbare, pe măsură ce valurile de erori se apropie de centrul său. Plasticitatea tot mai mare a nivelurilor sale mai periferice este unul dintre avantajele principale ale unui model predictiv *ierarhic*.

Al patrulea principiu, pe care nu l-am clarificat până acum, este că percepția (spre deosebire de învățare) *inversează direcția* prelucrării informațiilor. Prin inversarea dependențelor cauzale care au influențat inițial modelul predictiv, creierul produce inferențele noastre perceptive – pe care Merker le-a descris drept „o lume tridimensională, panoramică, bine definită, compusă din obiecte solide cu formă clară: lumea experienței noastre fenomenologice familiare”[4]. (Aceste inferențe circulă de la stânga spre dreapta în figura 15.) „Inversarea modelului predictiv” înseamnă pur și simplu trecerea de la învățare la predicție în baza a ceea ce ai învățat. E ceea ce făcea Eve Periaqueduct: infera („percepea”) starea lumii de dincolo de baraj din datele meteorologice și seismologice pe care le primea.

Percepția pornește *din interior spre exterior*, întotdeauna din punctul de vedere al subiectului. De fapt e vorba de apercepție, un proces deductiv, care ține de testarea ipotezei în paradigma bayesiană[5]. Hermann von Helmholtz, primul care a înțeles esența acestei chestiuni, a numit-o „inferență inconștientă” (din

nou, de notat adjectivul). Ceea ce vezi este „conjectura ta cea mai bună” cu privire la ce există în afara ta, e *propunerea ta de răspuns la întrebările pe care le adresezi acum lumii.*

Creierul trebuie să infereze cauzele cele mai probabile ale semnalelor de intrare fără să aibă acces direct la lumea imposibil de cunoscut aflată dincolo de zona lui. Creierul nu poate să o ia decât în direcția în care curg și se modifică stările sale senzoriale (*spike trains*; figura 11). Sarcina lui este să folosească aceste semnale ca să creeze un *model* probabilistic al regularităților ce există în lumea reală (sau, mai degrabă, între el însuși și lume), pe care le folosește apoi ca să genereze inferențe ce îi ghidează acțiunile – acțiuni care trebuie să-i asigure supraviețuirea în această lume. La rândul lor, acțiunile generează mostre senzoriale noi, ce sunt folosite pentru actualizarea în continuare a modelului, lucru pe care trebuie să-l facă pentru că modelele sunt imperfecte. Asta duce la alte acțiuni și așa mai departe.

Prin urmare, acțiunile trebuie privite ca *experimente* care testează ipoteze ce apar din modelul generativ. Dacă un experiment nu produce datele senzoriale preconizate, atunci sistemul fie (1) trebuie să-și schimbe predicția ca să explice datele mai bine, fie, dacă rămâne încrezător în predicția inițială, (2) trebuie să obțină date mai bune, adică să execute acțiuni care îi vor modifica inputul senzorial.

Aceste două opțiuni – schimbarea predicției sau a inputului – sunt mecanismele fundamentale ale percepției și, respectiv, acțiunii.

Ultimele trei paragrafe de mai sus mă determină să corectez un bias ce influențează acest capitol și capitolul anterior. Până

acum, ca majoritatea specialiștilor care folosesc funcțiile corticale drept model al felului de funcționare a creierului, m-am axat aproape exclusiv pe inferența *perceptivă* bayesiană. Dar există și inferență *activă*. De fapt, inferența activă este forma primară (cel puțin în biologie), întrucât percepția există ca să ghideze acțiunea.

După cum tocmai am precizat, creierul bayesian are două moduri de a răspunde la eroarea de predicție. Când se confruntă cu o ipoteză în cazul căreia se aplică o probabilitate „posterioară” în scădere, el creează o potrivire mai bună între ipoteză și date schimbându-și fie predicția „anterioară”, fie inputul. Diferența dintre aceste alternative se reduce la *direcția* statistică *a potrivirii*: eroarea este redusă dacă predicția este schimbată astfel încât să se potrivească inputului senzorial și este de asemenea redusă dacă inputul senzorial este schimbat astfel încât să se potrivească predicției. Bineînțeles, în realitate organismele pendulează tot timpul între aceste două opțiuni. (Să ne gândim la un șoarece-de-câmp care tot țopăie prin tufăriș, oprindu-se ca să privească în jur, țopăie din nou, se opreşte iar ca să privească și tot așa.) Din anumite puncte de vedere, percepția și acțiunea sunt mai asemănătoare decât par.

Corpul însuși este o lume ascunsă „exterioară” zonei Markov a sistemului tău nervos central. Valurile de predicții din figura 15 ar fi putut să includă straturi concentrice echivalente care să culmineze în receptorii și efectorii viscerali terminali care determină funcționarea organelor tale *interne*. Așadar, modelul despre lume al creierului trebuie să includă un model al sinelui tău corporal și al traiectoriei sale în aceeași măsură în care trebuie să includă un model al tuturor celorlalte cauze ascunse ce prezintă interes pentru tine. (În exemplul meu de mai sus,

nu ai fi putut ieși din încăperea plină cu fum dacă nu ai fi avut un model implicit al modului în care se mișcă corpul tău *și* în care îți echilibrează gazele sangvine.)

Mai mult, acțiunea nu apare deoarece creierul transmite vreun mare plan tuturor mușchilor și organelor corpului. În schimb, mușchii se contractă, iar glandele secretă până când dispar erorile de predicție pe care le transmit în ierarhie. Astfel, organele „acțiunii" musculoscheletale și viscerale ale corpului sunt la cheremul semnalelor de eroare generate prin diferențele dintre ce se așteaptă modelul predictiv să realizeze ele și ce realizează ele de fapt. Suprimarea erorii de predicție este ceea ce controlează acțiunea, nu mai puțin decât percepția[6].

Să ne amintim de legea lui Friston: toate cantitățile dintr-un sistem cu autoorganizare ce se pot schimba se vor schimba ca să minimizeze energia liberă. Numeroasele homeostate corporale reglate și orchestrate de metahomeostatul din mezencefal constituie pivotul mecanismului prin care ne menținem în viață, din simplul motiv că reglarea homeostatică ne ține corpul în limitele sale viabile. Aceste limite nu pot fi schimbate. Asta înseamnă – conform legii lui Friston – că *altceva din sistem trebuie să se schimbe*. Aceasta este explicația formală, mecanicistă a legăturii imperative dintre pulsiune și *acțiune*, precum și motivul pentru care trebuie să existe *o ierarhie a predicțiilor anterioare*, dintre care unele pot fi schimbate, iar altele nu[7].

Totuși, nu prea este de folos să acţionezi orbeşte. Acţiunea trebuie să fie ghidată de percepție, care este generată de un model al sinelui în lume. Regula lui Bayes descrie felul în care este implementat modelul predictiv ce realizează acest lucru,

modul în care se actualizează, astfel, constant, precum și motivul pentru care trebuie actualizat. Aceasta devine baza formală a *învățării*, prin care predicțiile sunt obținute și nuanțate în timp din semnalele de eroare primite. De asemenea, aceste dinamici ale sistemului confirmă, din rațiuni la fel de mecaniciste, că acțiunea are prioritate faţă de percepție: numai acțiunea poate să crească probabilitățile predicțiilor anterioare – dintre care unele, după cum am spus, pur și simplu nu pot fi schimbate.

Regula lui Bayes pleacă de la o asumpție a ceea ce am numit „cunoaștere de fond”. Altfel, ea nu poate funcționa[8]. Apare astfel o întrebare: de unde vine cunoașterea de fond, la început, înainte ca sistemul să adune dovezi despre lume? Răspunsul este că principalele noastre „stări prognozate” sunt codificate de specia noastră ca puncte de stabilizare homeostatice *înnăscute* – cantități care au fost determinate de ceea ce a funcționat efectiv pentru strămoșii noștri pe parcursul evoluției. Noi suntem beneficiarii succeselor biologice ale generațiilor trecute, care stabilesc premisele elementare ale existenței noastre.

Și nu ne putem culca pe laurii lor. Legătura dintre afect și acțiune dictează că, dacă nu reușești de la început, trebuie să tot încerci. Cererile celor mai profunde pulsiuni biologice ale tale sunt inexorabile: intensitatea lor scade numai dacă le satisfaci sau dacă mori. Cu excepţia celei de-a doua situații, trebuie să-ți suplimentezi reflexele și instinctele înnăscute și să dezvolți *alte* modalități de a-ţi satisface nevoile. Nu există alternativă. Cu alte cuvinte, trebuie să înveți din experiență. Din fericire pentru noi, creierul uman este neobișnuit de bine echipat pentru asta.

O implicație interesantă a tuturor acestor lucruri este că, dacă afectul chiar funcționează conform mecanismelor homeostatice pe care le-am descris – dacă într-adevăr este „o măsură a cererii de travaliu impuse psihicului ca urmare a legăturii sale cu ceea ce e corporal” –, atunci el trebuie să fie *adevăratul* vehicul fundamental al minimizării energiei libere. Așadar, afectul este mediul primar al volițiunii – și izvorul întregii vieți psihice.

Am spus mai sus că învățarea din experiență produce un model ierarhic al lumii care, atunci când este inversat, generează predicții despre aceeași lume. Dar procesul nu se încheie aici – predicțiile trebuie *testate*. De aici apar erori de predicție, care sunt folosite pentru actualizarea modelului. Asta înseamnă învățare „din experiență”. Erorile de predicție sunt semnalele senzoriale care *nu* au fost prezise de o ipoteză curentă, deci cele care nu au fost autogenerate. Aceasta este partea esenţială a datelor.

Greșeala pe care o fac în acest punct majoritatea specialiștilor în științe cognitive este să presupună că datele de intrare sunt exclusiv *exteroceptive*. Ei uită că erorile de predicție (inputurile senzoriale) care contează cel mai mult pentru noi vin din *interior*. De exemplu, devierea de la temperatura corporală preconizată furnizează feedback „senzorial”, nu mai puțin decât evenimentele exterioare nepreconizate; la fel și semnalul de eroare homeostatică ce declanșează alarma de sufocare. Aceste semnale generează *afecte*, și nu percepții. După cum a afirmat Freud, prozencefalul este un „ganglion simpatic”. Confuzia din acest punct este prețul etern pe care îl plătesc colegii mei pentru că acceptă eroarea privitoare la cortex[9]. Conștiința este generată la nivel

endogen – toată. Conștiința, la origine, este afect. Apoi este extinsă spre exterior către percepție, ca să evalueze inferențele perceptive, în modul pe care îl voi descrie în continuare.

În sfârșit, ne putem ocupa de întrebarea: de ce și cum anume *devin conștiente* funcțiile naturale de autoconservare descrise în acest capitol? Știm că conștiința e înrădăcinată în afecte, în sentimente. Dar care sunt legile formale, mecaniciste care dau naștere sentimentelor – și, astfel, conștiinței?

9

De ce și cum se naște conștiința

Întrebarea fundamentală pe care trebuie să și-o pună mereu ființele vii este: „Ce se întâmplă cu energia mea liberă dacă fac asta?”. Dar să faci ce? Direcțiile potențiale de acțiune în orice moment nu sunt arbitrare sau infinite, ci sunt dictate de *nevoile* curente. Există, desigur, o legătură strânsă între nevoi și acțiuni – fiecare nevoie cere o anumită acțiune adecvată. Dacă îți este foame, trebuie să mănânci. Dacă ești obosit, trebuie să te odihnești. Totuși, există obstacole în calea execuției: nu poți face decât unu sau două lucruri în același timp. Asta înseamnă că, pentru a-ți selecta următoarea acțiune, trebuie să-ți ordonezi nevoile curente în funcție de urgență.

Există două lucruri care fac sarcina ordonării mai complicată decât ar putea să pară la prima vedere. Primul: de obicei, nu e necesar ca nevoile noastre, de organisme complexe, să fie satisfăcute într-o ordine fixă. Să mănânci sau să dormi – ce e mai important? Depinde de tot felul de considerații. Al doilea: multe dintre nevoile organismelor complexe nu pot fi satisfăcute întotdeauna prin aceeași acțiune. Ca să mănânci supă, îți trebuie abilități diferite de cele necesare ca să mănânci porumb fiert, de exemplu (și nici nu mă gândesc la abilitățile și

resursele implicate în prepararea lor). În ambele cazuri, „ce faci în continuare” depinde neapărat de context. De aceea, nevoile interioare trebuie prioritizate în funcție de *condițiile exterioare predominante*.

Ne naștem având predicții specifice speciei despre ce e de făcut în stări precum foamea, setea, frica și furia. Aceste predicții înnăscute sunt numite „reflexe” și „instincte” – instrumente de supraviețuire moștenite pentru care avem toate motivele să fim recunoscători. Dar ele nu sunt suficient de flexibile ca să facă față diversității și complexității situațiilor cu care ne confruntăm de fapt, așa că trebuie completate. Acesta este rolul învățării din experiență.

Am arătat deja legătura dintre învățarea din experiență și legea afectului. Valența afectivă – sentimentele noastre despre ce este „bun” și „rău” din punct de vedere biologic pentru noi – ne ghidează în situații *neprevăzute*. Am concluzionat că acest fel de a ne găsi drumul printre problemele neprevăzute ale vieții în funcţie de ceea ce simţim, cu ajutorul comportamentului *voluntar*, este funcția biologică a conștiinței. Ea ne ghidează alegerile când suntem în întuneric. Dar, desigur, ca să poată face acest lucru, ea trebuie să creeze legătura dintre afectele noastre interioare (înrădăcinate în nevoi) și reprezentările lumii exterioare.

Acest lucru explică de ce activarea este însoțită de sentimente *și* de percepții conștiente ale lucrurilor. Am recunoscut la sfârșitul capitolului 6 că marele mister al acestei legături – misterul modului în care experiența subiectivă se potrivește în materia universului fizic – poate fi elucidat numai dacă reducem atât fenomenele fiziologice, cât și cele psihologice la cauzele lor mecaniciste fundamentale. Aceste

cauze urmau să fie dezvăluite la un nivel de abstracție pe care numai *fizica* îl putea furniza. În ultimele două capitole am demarat o descriere formală a acestor mecanisme unificatoare.

Acum a sosit timpul să finalizez descrierea. Dacă autoorganizarea și homeostaza nu explică în sine de ce și cum apare conștiința, atunci care este explicația? Care este legătura, formal și mecanic, dintre procesul biologic al *prioritizării nevoilor* pe care tocmai l-am rezumat și minimizarea energiei libere? Și cum se face că rezultatul acestui proces *este simțit ca ceva* de către și pentru unele sisteme cu autoorganizare?

Punctul de plecare al răspunsului meu este exact ceea ce tocmai am evidențiat: creaturile complexe ca noi, vertebratele, au multe tipuri de nevoi. Adică avem multe sisteme interne, fiecare reglat de propriile mecanisme homeostatice, toate contribuind cu valori de eroare la calculul total al energiei libere. Nevoile noastre biologice sunt aceste valori de eroare. Când nevoile sunt resimțite ca afecte, le descriem ca având „valențe” pozitive sau negative. Asta înseamnă că au o *valoare* subiectivă: ni se par bune sau rele pentru noi. Behavioriștii au încercat să obiectiveze valoarea redefinind sentimentele plăcute și neplăcute drept stimuli care recompensează și care pedepsesc – dar ne-am ocupat deja de această chestiune. Valența nu ține de stimul, ci este, inerent, subiectivă și calitativă. Ceva poate fi încântător pentru o persoană și înfricoșător pentru alta.

Dar este oare *imposibil* să cuantificăm valența? Să analizăm figura 12: cu cât săgeata deviază spre dreapta, cu atât mai mare este neplăcerea. Astfel, în orice moment, valoarea foamei tale ar putea fi 3/10 (care e mai rea decât 1/10), iar valoarea setei

tale ar putea fi 2/10 (care e mai bună decât 5/10) etc. Specialiștii în afecte fac tot timpul asemenea măsurători, le cer participanților la cercetare să-și evalueze plăcerile și neplăcerile pe ceea ce se numește scară Likert. Deși aceste scări sunt subiective, faptul că sunt cuantificabile *în principiu* lasă deschis răspunsul la întrebarea: de ce trebuie calificate afectele? Dacă sistemele cu autoorganizare pot să consemneze „răspunsurile evocate de echipament" (adică propriile stări) drept cantități, în principiu, atunci care este contribuția calității? Această întrebare se referă la ceea ce filosofii numesc „qualia" – materia minții care ne scapă și despre care se spune că nu i se poate găsi locul în concepția noastră fizicalistă asupra universului.

Răspunsul începe de la faptul că nevoile nu pot fi combinate și adunate într-un mod simplu. *Multiplele noastre nevoi nu pot fi reduse la un singur numitor comun,* ci trebuie să fie evaluate pe scări separate, aproximativ egale, astfel încât fiecare să fie recunoscută. Nu poți spune pur și simplu că „3/10 din foame plus 1/10 din sete este egal cu 4/20 din nevoia totală", apoi să încerci să minimizezi valoarea totală, deoarece fiecare nevoie în parte trebuie satisfăcută. Metabolismul energetic nu este același lucru cu hidratarea, care nu este același lucru cu termoreglarea și așa mai departe; fiecare este esențială. Conform lui Edmund Rolls, specialist în neuroștiințe și behaviorist, „dacă recompensa alimentară ar fi întotdeauna mult mai puternică decât alte recompense, atunci genele animalului nu ar supraviețui, pentru că el nu ar bea niciodată apă"[1].

Dacă analizăm acești factori laolaltă, are sens ca sistemele biologice cu autoorganizare să își diferențieze nevoile (valorile

de eroare) pe criterii *categoriale*. Deosebirea dintre variabilele categoriale este *calitativă*. Întrucât nu se poate pune semnul egalității între tipul de eroare *A* de 8/10 și tipul de eroare *B* cu aceeași valoare, din motivele pe care tocmai le-am explicat, ele trebuie tratate ca variabile categoriale. Astfel, sistemul poate să o recunoască pe fiecare în parte pe termen lung *și* să le prioritizeze în funcție de context. De aceea, are sens ca sistemele complexe care se autovalidează să-și clasifice numeroasele homeostate, astfel încât să le poată estima independent unul de celălalt și să prioritizeze efectele.

Nu numai că nevoile diferite contribuie cu cantități diferite la energia liberă totală, ci și cantitățile diferite au implicații diferite pentru animal în contexte diferite (de exemplu, foamea trece înaintea somnolenței în unele situații, dar nu şi în altele). Acest lucru contribuie semnificativ la incertitudine – duşmanul de moarte al mecanismelor predicției. Incertitudinea tot mai mare este o situație periculoasă pentru orice sistem cu autoorganizare: prezice moartea sistemului. Mai multă incertitudine cere mai multă complexitate de calcul (adică mai mult flux informațional, adică mai multă entropie). Astfel, clasificarea devine o necesitate atunci când valoarea relativă a unor cantități diferite *se schimbă* în timp (dacă 8/10 pentru *A* valorează acum, dar nu întotdeauna mai mult decât 8/10 pentru *B*).

Ne putem imagina că s-ar putea dezvolta un set extrem de complex de algoritmi ai modelului pentru a calcula cereri relative de supraviețuire în toate situațiile predictibile, pentru a ne permite *automat* să prioritizăm acțiunile pe acest criteriu. Totuși, asemenea modele complexe sunt extrem de costisitoare, în toate sensurile cuvântului. Sunt greu de gestionat, ceea ce

înseamnă întârziere, ce poate face diferența dintre viață și moarte; necesită o mare capacitate de procesare, ceea ce înseamnă că trebuie să găsească mai multe resurse energetice. Statisticienii numesc „explozie combinatorie” creșterea exponențială a resurselor de calcul ce sunt necesare unei creșteri liniare a complexității modelului.

Mai mult, un model complex care prezice cu acuratețe ce se întâmplă într-o situație specifică probabil că nu va prezice cu aceeași acuratețe ce se va întâmpla în alte situații. În termeni statistici, spunem că modelele excesiv de complexe „supraajustează” un eșantion de date. Modelul fisurilor preconizate în baraj alcătuit de Eve Periaqueduct nu se baza pe evenimente monitorizate oră de oră sau zi de zi în cele câteva săptămâni anterioare, ci pe niște medii pe termen lung, calculate pe baza unui număr mare de eșantioane de date culese de-a lungul mai multor ani. Acest lucru i-a făcut modelul mai simplu și, de aceea, mai *generalizabil*. Conform „briciului lui Occam” (principiul parcimoniei)[2], vrem modele predictive *simple*. Simplificarea este esențială pentru ca modelele noastre să se aplice într-o mare varietate de situații. Trebuie să fie utile, nu numai aici și acum, ci și în multe alte contexte.

Așadar, modelele predictive trebuie să fie simple. Însă, după cum a spus Einstein, „totul trebuie făcut cât de simplu se poate, dar nu mai simplu”[3]. Cum ajungem la echilibrul corect? Compartimentarea este metoda statistică standard folosită pentru a obține echilibrul optim între complexitate și acuratețe. Asta se poate face în mai multe feluri. De exemplu, o parte din creierul vizual calculează *ceea ce* vezi, iar o alta calculează *unde* se află acel ceva. Astfel, ai posibilitatea să presupui identitatea constantă a ceva în timp ce se mișcă în jurul tău, își schimbă

forma, mărimea și orientarea. Dar cel mai mult contează că o capacitate de a compartimenta permite sistemului să-și ordoneze nevoile și predicțiile lor aferente (adică sursele predominante ale energiei libere preconizate) în categorii, de-a lungul timpului, și să-și concentreze eforturile de calcul asupra compartimentului prioritizat.

Aceasta este baza statistico-mecanică a realității că fiecare afect are nu numai o valență hedonică *continuă* (un *grad* de plăcere și de neplăcere – aspect comun tuturor afectelor), ci și o *calitate categorială* (astfel încât, de exemplu, setea se simte diferit de detresa de separare, care se simte diferit de dezgust și așa mai departe). Acestea sunt caracteristici esențiale ale qualia afective, forma elementară a tuturor qualia: ele au atât cantitate, *cât și calitate*. Mai exact: afectele sunt întotdeauna subiective, calitative și au valențe. Ele trebuie să fie așa, întrucât au evoluat astfel încât să rezolve problema controlului.

E util să descriem selectarea și prioritizarea în funcție de categorii ale afectelor în termeni de modelele operaționale. Să ne gândim la felul cum se comportă un avion când este în modul „decolare” versus modul „croazieră” versus modul „turbulențe” versus modul „aterizare”. În aceste situații diferite sunt implicate aceleași variabile, dar ele trebuie *cântărite* diferit de fiecare dată. De exemplu, altitudinea exactă contează mult mai mult în timpul aterizării decât în timpul regimului de zbor de croazieră. Același lucru se aplică în cazul barajului de care se ocupa Eve Periaqueduct. În modul „iarnă” versus modul „vară” și în modul „cutremur” versus modul „noncutremur” se aplicau programe de funcționare diferite. În condiții de cutremur, Eve ar fi obligată să ignore programul sezonier normal, automatizat, ca să implementeze un program de tip

„urgență de cutremur”.

În termenii fiziologici pe care i-am folosit în capitolul 6, diferitele moduri de operare sunt funcții *de stare* ale creierului. Așa cum am explicat, triunghiul de decizie din mezencefal selectează stări cerebrale afective – precum modul „alarmă de sufocare”. Asta se întâmplă când PAG, substanța cenușie periapeductală, răspunde la întrebarea: „Care dintre aceste semnale de eroare convergente (adică nevoi) constituie cea mai mare ocazie de a-mi minimiza energia liberă?”. Cu alte cuvinte, care nevoie este predominantă *în acest moment*? Răspunsul vine nu numai din amploarea relativă a semnalelor de eroare concurente, ci și din diferențele dintre categorii (moduri sau stări), a căror predominanță trebuie evaluată în context. După cum am explicat, informațiile contextuale sunt furnizate de colicului superiori.

Iată un exemplu pe care l-am observat astăzi. Când am ieșit să fac jogging, la 7 dimineața, era întuneric și, când m-am întors acasă, o oră mai târziu, era lumină. (E iarnă şi locuiesc într-o zonă rurală din Sussex, unde scriu această carte.) La plecare, am trecut pe lângă un câmp aflat în apropierea unei ferme, unde o turmă de oi m-a observat și numai nu s-au călcat în picioare înghesuindu-se să fugă de mine. La întoarcere, când am trecut pe lângă același câmp, aceleași oi, aflate în același loc, abia dacă s-au uitat la mine. Spaima lor în contextul întunericului a fost înlocuită de plictiseală în timpul zilei. Pe scurt, contextul a modificat semnificația evenimentului „om care aleargă spre mine”. Noaptea, acest eveniment este prioritizat, fapt care introduce brusc oile în modul FRICĂ, dar nu și ziua, așa că oile rămân în modul implicit CĂUTARE.

Astfel se determină ce face sistemul în continuare. Cu alte

cuvinte, se determină stările *active* care vor fi selectate de modelul generativ pentru a rezolva categoria de incertitudine ce a fost prioritizată. E ca și cum sistemul spune: în condițiile actuale, *aceasta este categoria de procesare a erorii de predicție în care nu poate fi sacrificată complexitatea de calcul.* Așa că sistemul (în acest caz, oaia) trece în modul de operare FRICĂ. Aplică strategia cea mai bună pe care o poate furniza modelul său generativ în circumstanțele date: fuge. Apoi, după ce a luat în considerare tot ce a învățat despre câmpul pe care se află (contextul preconizat), speră la ce e mai bine, dar se pregătește pentru ce e mai rău. Adică își execută planul cu prudență, gata să se adapteze la evoluțiile în desfășurare.

Este esențial că trecerea în modul FRICĂ înseamnă că nevoia prioritizată a devenit un afect. Cu alte cuvinte, a devenit conștientă. De ce? Devine conștientă pentru ca *devierile de la rezultatele preconizate* în categoria de nevoi predominantă să fie simțite în întreaga ierarhie predictivă. Iată ce este afectul. E „răspunsul evocat de echipament” la întrebarea despre sine pe care și-a pus-o sistemul: „Care dintre aceste semnale de eroare convergente constituie cea mai mare ocazie de a-mi minimiza energia liberă?”.

În Africa de Sud, unde locuiesc în cea mai mare parte a timpului, ai nenumărate ocazii să vezi cum funcționează în condiții naturale selectarea afectului, adică în tipurile de condiții în care a evoluat inițial acest mecanism. Mă refer la ce se întâmplă nu între lei și gazele în rezervaţiile noastre naturale fabuloase (cu toate că și aici găsești multe ocazii), ci în societatea noastră foarte inegală și deci afectată de multe probleme. Mulți dintre compatrioții mei știu cum e când nevoia ta predominantă este să scapi de cineva care încearcă să te

omoare. În acest punct, comportamentul tău voluntar devine dominat de sentimente de FRICĂ, ce evaluează succesul sau eșecul aici-și-acum al acțiunilor tale aflate în desfășurare. Alte nevoi (precum nevoia de a urina) sunt pasate modului automat. Cu alte cuvinte, e posibil să faci pe tine fără ezitare.

Încercăm să reducem fenomenele fiziologice și psihologice ale activării afective la un set de principii mecanice care pot fi formalizate matematic. În capitolul 6 am explicat că, odată ce triunghiul de decizie din mezencefal a prioritizat o nevoie, modelul prozencefalului despre sinele său în lume generează un context preconizat în care nevoia respectivă va fi satisfăcută. Am spus că această lume preconizată are două fațete: pe de o parte, ea reprezintă conținutul real al predicțiilor noastre, iar pe de altă parte, trebuie să codifice nivelul nostru de încredere în predicții. Prima dintre aceste fațete este furnizată de rețelele memoriei pe termen lung din prozencefal, care filtrează prezentul prin prisma trecutului. Am prezentat în capitolul 7 principiile care le guvernează. A doua fațetă – ajustarea nivelurilor de încredere – este modulată de „activare". Așadar, să formalizăm legile care o guvernează.

Primul mecanism identificat până acum este cel prin care este selectată categoria predominantă de acțiune (cea mai eficientă modalitate sau stare de funcționare). Acesta este modul în care anumite calități afective mai întâi reglează acțiunile unor sisteme complexe cu autoorganizare. Asta duce la generarea unor planuri senzorimotoare prin care sistemul „speră la ce e mai bine, dar se pregătește pentru ce e mai rău". Iată detaliile asupra cărora vreau să zăbovesc acum. Care este mecanismul cauzal prin care se reglează speranța la ce e mai

bine și pregătirea pentru ce e mai rău?

Prima parte a răspunsului este că nivelurile de încredere atașate predicțiilor sunt *învățate din experiență*, la fel ca orice altceva. Apoi, nivelul de încredere în predicțiile noastre poate fi anticipat, ca și predicțiile. Codificarea predictivă ne impune să asociem probabilități stărilor senzoriale care preconizăm că vor decurge din anumite acțiuni, iar apoi să comparăm probabilitățile cu distribuțiile observate concret în eșantioanele senzoriale ce urmează. Aceasta este esența metodei bayesiene de „actualizare a ipotezelor” pe care am descris-o mai sus – metoda prin care ne minimizăm energia liberă.

Acum, ca să stabilim potrivirea dintre un model și unele date, nu e suficient să comparăm pur și simplu *media* distribuțiilor, ci e necesar și să evaluăm şi *variația mediei* (vezi figura 16). O valoare mare a variației într-un eșantion reduce încrederea în potrivire. Dacă o știre spune „A murit regele… A murit regele… A murit regele”, e mai probabil să o iau în serios decât dacă spune „A murit regele… nu, regele nu a murit… de fapt, se pare că regele s-ar putea să fi murit până la urmă”. Evaluările cu privire la diferența dintre o distribuție prevăzută și un eșantion de date sunt mai ușor de făcut atunci când distribuția este îngustă și *precisă*.

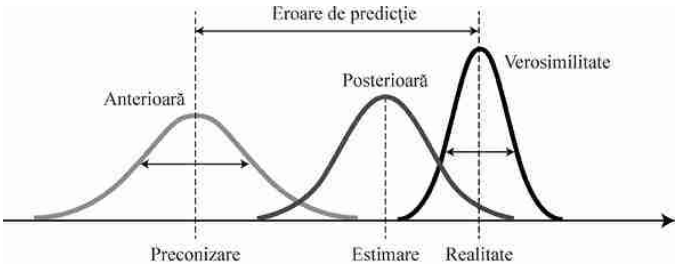


Figura 16. O predicție anterioară (stânga) este comparată cu un

eșantion de date senzoriale (dreapta), rezultând o predicție posterioară (mijloc). „Mediile” acestor trei distribuții sunt indicate de liniile verticale (punctate), iar „variația” lor este indicată de liniile orizontale (cu săgeți). Variația preconizată mare în distribuția anterioară (linia orizontală din stânga) indică un grad scăzut de încredere în predicția anterioară, iar variația reală îngustă din eșantionul de date senzoriale (linia orizontală din dreapta) indică un grad ridicat de încredere în date. În acest exemplu, precizia (variația inversă) a datelor senzoriale este mare; în consecință, predicția posterioară trece decisiv spre dreapta. Dacă precizia datelor senzoriale ar fi mai mică, predicția posterioară s-ar deplasa mai puțin sau chiar deloc.

Obiectivul nostru în interacțiunile cu lumea trebuie să fie *precizia*. De aceea, modelele noastre trebuie să aibă un mecanism pentru *predicția preciziei*. Asta ne permite „să cântărim” precizia preconizată a semnalelor de eroare primite în ce privește precizia pe care am atribuit-o unei predicții de ieșire. Acest lucru (gradele relative de încredere) va dicta influența semnalelor de eroare reale asupra predicțiilor noastre. Dacă avem tot mai multă încredere într-un semnal de eroare primit, ar trebui să avem tot mai puțină încredere în planul nostru curent de acțiune, iar evoluțiile nepreconizate vagi ar trebui să nu ne îndepărteze de la cursul nostru predeterminat. (Se pot atribui valori de încredere relativă și așteptărilor active versus așteptările perceptive și celor exteroceptive versus cele interoceptive, precum și tuturor celorlalte cantități implicate în legea lui Friston[4].) Aceasta este

un fel de inferență bayesiană de ordinul doi, care atrage după sine *inferențe despre inferențe*, adică niveluri de încredere informată despre predicții.

Scopul modulării preciziei este de a garanta că inferențele făcute de modelele predictive sunt influențate de semnale de învățare *fiabile* (știri demne de încredere): dacă există o încredere mare într-un semnal, atunci trebuie să i se permită să revizuiască o ipoteză anterioară și invers dacă încrederea e mică. Încrederea afectează forța semnalelor de eroare ce sunt propagate spre interior prin ierarhie. Un semnal în care ai încredere mare (adică unul mai precis) este mai „sonor”. Prin urmare, va avea o șansă mai mare să transmită o eroare reziduală spre centrul sistemului și să-și actualizeze modelul generativ. În schimb, semnalele mai puțin precise – în care ai mai puțină încredere, cunoscute și ca „zgomot” – pot fi izolate în epiteliul senzorial și ignorate fără riscuri (așa sperăm).

Asta înseamnă că trebuie să minimizăm semnalele de eroare precise. Din nou, pare paradoxal, până ce realizezi că nu e vorba decât de ideea că trebuie să evităm să facem greșeli monumentale. Asta nu se poate realiza decât prin îmbunătățirea modelelor noastre generative, crescând astfel informarea reciprocă între modelele noastre despre lume și eșantioanele senzoriale obținute din ea. Cu alte cuvinte, trebuie să maximizăm precizia predicțiilor noastre, apoi să căutăm date precise care le confirmă. *Trebuie să ne maximizăm încrederea în credințele care ne ghidează acțiunile*, ceea ce se numește „optimizarea preciziei”.

Realizăm optimizarea preciziei învățând din experiență. Trebuie să învățăm în ce surse de informare putem avea încredere (și când) și să ne ajustăm predicțiile în consecință. De

exemplu, avem încredere în semnalele vizuale ziua și în semnalele auditive noaptea. Acordăm o pondere mai mare celor văzute ziua și celor auzite noaptea, deoarece *am învățat* să procedăm astfel. În consecință, ne așteptăm ca informațiile vizuale să fie mai precise ziua decât noaptea. Acest lucru se numește „precizie preconizată”. Pentru că nu ne așteptăm la precizie vizuală noaptea, tolerăm fără ezitare tot felul de imagini neclare și vagi, dar niște experiențe vizuale similare în timpul zilei ne-ar face să credem că ceva nu e deloc în regulă.

La fel, angajaţii lui Eve Periaqueduct au învățat că presiunea în scădere în barometre prezice creșterea precipitațiilor iarna, dar nu şi vara. Prin urmare, ei puteau să considere „zgomot” o valoare în scădere cu mai multă încredere vara decât iarna. Se *așteptau* ca valorile barometrelor să fie mai puțin precise vara. În caz contrar, își ajustau nivelurile de precizie preconizate, care, apoi, le afectau predicțiile posterioare. Să dau un exemplu mai afectiv: oile din Sussex au învățat să aibă încredere în predicția lor că oamenii care aleargă spre ele ziua nu le fac rău, dar au mai puțină încredere în aceeași predicție noaptea. În consecință, îi atribuie mai puțină precizie preconizată în timpul nopții. Poate că, dacă vor continua să dea de mine în condiții de întuneric și se va dovedi că nu le fac niciun rău, își vor ajusta această valoare a preciziei, modificându-și în consecință predicțiile despre oamenii care aleargă spre ele noaptea. Chestiunile dependente de context pot fi învățate, ca orice altceva.

E clar din afirmațiile mele că acest tip de învățare se învârte în jurul *contextelor fluctuante*. În lipsa modelării predictive a dinamicii contextuale, un sistem cu autoorganizare nu poate supraviețui mult timp în mediile schimbătoare. Modelul

generativ trebuie să încorporeze aceste dinamici. Trebuie să învețe să anticipeze grade de precizie. Iar ajustarea valorilor preciziei, ca orice altceva în creierul predictiv, trebuie să respecte legea lui Friston.

Precizia este modul în care creierul își reprezintă gradul de încredere într-o anumită sursă de dovezi senzoriale sau în consecințele preconizate ale unei anumite acțiuni. Valorile preciziei cuantifică așteptările în ceea ce priveşte variabilitatea. Așadar, ele sunt *reprezentări ale incertitudinii*. Câtă încredere am în semnalul de eroare *x* în contextul curent? Câtă *pondere* să-i dau, chiar acum? 8/10 pentru *A* valorează mai mult sau mai puțin decât 8/10 pentru *B în condițiile din acest moment*?

Deja am văzut fiziologia acestui lucru: triunghiul de decizie din mezencefal prioritizează o nevoie, apoi modelul sinelui în lume din prozencefal generează un context preconizat în care nevoia prioritizată va fi satisfăcută. Lumea preconizată are două fațete: conținutul efectiv al predicțiilor și nivelul de încredere a sistemului în predicțiile respective. Acum, că știm cum sunt cuantificate nivelurile de încredere, le putem insera în descrierea pe care am făcut-o fiziologiei activării.

După ce triunghiul de decizie și-a selectat nevoia predominantă curentă, ceea ce determină starea afectivă a sistemului, care determină apoi contextul preconizat generat de sistemele de memorie pe termen lung din prozencefal, sistemul reticulat activator se pune pe treabă. Sistemele mnezice atribuie *valori de referinţă ale preciziei* pentru contextul preconizat și le aplică în toată ierarhia predictivă. Apoi un nor de neuromodulatori se vălătucește în prozencefal, impulsionând unele canale să se descarce rapid și descurajând altele. Aceste frecvenţe de descărcare determină câtă pondere

se va atribui predicțiilor curente și erorilor lor aferente, ceea ce determină cât de „sonor” se vor transmite erorile. Cu alte cuvinte, valorile preciziei determină *nivelul de încredere* al sistemului în legătură cu efectele care preconizează că vor decurge din cursul acțiunii aflate acum în desfășurare, la diversele niveluri ale ierarhiei. Apoi, din nou, speră la ce e mai bine și se pregătește pentru ce e mai rău.

Multe lucruri surprinzătoare pot fi prezise odată ce te obișnuiești cu ele. Totuși, dacă nu putem decât să sperăm la ce e mai bine și să ne pregătim pentru ce e mai rău, asta înseamnă că unele lucruri *nu pot* fi prezise. Iar aceasta este a doua parte a poveștii. Asta înseamnă că sistemul trebuie să-și ajusteze nivelurile de încredere pe moment – adică *să moduleze* activarea în contextul evenimentelor aflate în desfășurare, pe măsură ce ele se întâmplă.

Am afirmat mai sus că ne putem imagina că se dezvoltă un set extrem de complex de algoritmi ai modelului (indiferent cât de complicați devin aceștia), care calculează cererile relative de supraviețuire în toate situațiile predictibile, și că acesta își prioritizează opțiunile de acțiune în baza acestor calcule, indiferent de „explozia combinatorie”. Dar cum alege organismul între *A* și *B* când *incertitudinea în sine* devine elementul principal ce determină selectarea acțiunii? Asta se întâmplă în situațiile noi, de exemplu, care nu sunt în niciun caz rare în natură.

Fiziologii numesc „modulare a activării” ceea ce specialiștii în științe computaționale numesc „evaluarea preciziei”[5]. Sunt același lucru. După cum tocmai ai văzut, un semnal de precizie nu este altfel decât ceea ce am numit semnal „sonor” în

capitolul 6 – unul *puternic*. Asta presupune că modularea încrederii într-un semnal de eroare trebuie să urmeze *deviațiile de la forța lui preconizată*. Iar aceste deviații trebuie minimizate. Ca în cazul tuturor semnalelor de eroare homeostatice, e „bine” (pentru noi, sistemele biologice) când lucrurile se întâmplă după cum am prezis și „rău” când predomină incertitudinea.

Pe măsură ce se desfășoară secvența de acțiune, nivelurile de încredere inițiale sunt *ajustate* ascendent și descendent de sistemul reticulat activator. (Să ne amintim de angajații lui Eve Periaqueduct care citesc și ajustează instrumentele.) Cu alte cuvinte, contextul senzorimotor în desfășurare este „palpat” – și evaluările încrederii sistemului ajustate – pe baza fluctuațiilor în desfășurare în incertitudinea preconizată. Modificările la nivelul activării urmăresc coeficientul de încredere estimată al erorilor de predicție testate. În acest mod, valorile de precizie fluctuante estimează coeficientul de încredere schimbător al semnalelor aflate în desfășurare ce conţin noutățile. La rândul lor, aceste valori determină tot ceea ce mai face sistemul, conform legii lui Friston.

Toate acestea sugerează că *optimizarea preciziei este baza statistico-mecanică a prioritizării semnalului în general* – adică outputul critic a tot ce am văzut că se petrece în triunghiul de decizie din mezencefal și în sistemul reticulat activator. Optimizarea preciziei este modul în care numeroasele semnale de eroare ce converg în PAG au fost prioritizate de la început, aducând în conștientizarea afectivă nevoia predominantă, ducând la o serie de opțiuni în curs de derulare într-un context preconizat, ghidat de precizii preconizate. Acum, acestea trebuie să fie modulate ținând cont de evenimentele senzoriale

nepreconizate.

Poate că toate acestea sună mai degrabă abstract. Din contră, eu cred că sunt descrieri a ceea ce se întâmplă concret. Atât de mult din experiența noastră constă pur și simplu din mici impulsuri senzoriale, când observi lucruri care nu sunt chiar cum te-ai aștepta să fie, urmate de sondări cognitive după moduri în care să elimini diferența. Îți amintești că trebuie să trimiți un e-mail: abia când mâna ta nu dă de ecranul telefonului îți dai seama că deja îl căutai – dar, dacă telefonul nu e chiar acolo, lângă tine, atunci unde l-ai lăsat? În bucătărie, unde erai acum cinci minute?

Sau să luăm un exemplu mai semnificativ din punct de vedere afectiv: o întâlnire cu un posibil partener romantic. Crezi că l-ai putea (ai putea-o) cuceri diseară. Îți imaginezi succesiunea posibilă a evenimentelor și îți stabilești un plan de acțiune. Apoi speri la ce e mai bun. Nu ești sigur ce-o să iasă, dar, în baza experienței tale anterioare cu această persoană, îți evaluezi șansele la aproximativ 7/10.

Pe parcursul serii, aspectele pe care te axezi (ceea ce este predominant) sunt destul de diferite de cele pe care te-ai fi axat dacă ai fi luat cina cu fratele tău. Și tonul emoțional este diferit. Fiecare mic semn care sugerează că persoana aflată de cealaltă parte a mesei răspunde pozitiv la propunerile tale îți provoacă valuri de entuziasm: tot mai multă încredere că planul tău funcționează. Pe neașteptate, partenerul la care speri cască și se uită la ceas. Ce înseamnă asta? Câtă pondere să atribui acestei evoluții? Îți fuge pământul de sub picioare. Oare ai interpretat greșit toate semnele anterioare? Analizezi fiecare mișcare și gest. Cel mai mic indiciu suplimentar că sentimentele tale nu sunt împărtășite te va pregăti pentru ce e mai rău și vei iniția

planul B: îți salvezi mândria prefăcându-te la fel de indiferent. Apoi privirile vi se intersectează. Oare asta înseamnă ce crezi tu că înseamnă? Da, așa e! Apoi mâna îți este atinsă ușor. Inima începe să-ți bată mai tare. Se pare că, până la urmă, poți să rămâi la planul A.

O nevoie prioritizată (în acest caz, DORINȚA SEXUALĂ) este sursa de incertitudine predominantă în acest moment. Inferențele despre cauzele ei devin conștiente ca afect, deoarece fluctuațiile nivelului tău de încredere referitoare la posibilele acțiuni necesare pentru a satisface această nevoie trebuie să fie modulate de sentimente. Sentimentele îți spun cât de bine sau cât de prost te descurci. Contextul în desfășurare ce generează fluctuațiile trebuie să devină conștient de asemenea, din același motiv. Iată de ce am definit conștientizarea exteroceptivă a acțiunii și percepției drept *afect contextualizat*. Acum avem o înţelegere formală, mecanicistă a ceea ce înseamnă asta. Totul nu e decât incertitudine resimțită.

Este crucial să observi că afirmația „Contextul în desfășurare ce generează fluctuațiile trebuie să devină conștient” explică de ce experiența are aspecte duale. Nu e doar o chestiune de „mă simt așa”, ci de „mă simt așa *în legătură cu asta*”. Și „în legătură cu asta” trebuie resimțit folosind o monedă comună (incertitudinea aplicată) – deoarece contextul este principala sursă a incertitudinii referitoare la energia liberă. Economia minimizării energiei libere necesită o monedă comună.

Aceste fapte relevă că conștiința nu este doar o perspectivă subiectivă asupra dinamicii „reale” a sistemelor cu autoorganizare, ci şi o funcție având propriile puteri cauzale determinate. *Sentimentul* unei nevoi (spre deosebire de simpla

existență a unei nevoi) contează enorm pentru ce va face în continuare subiectul nevoii respective. Afectele orientează efectiv ce face un animal de la un moment la altul în condiții de incertitudine. Percepțiile exteroceptive există exclusiv pentru că sunt resimțite *în relație cu* acțiunile orientate afectiv pe care le contextualizează.

Aceasta este funcția centrală a unor lucruri precum atenția. Concentrarea atenției funcționează ca selectarea afectului, dar este aplicată lumii exterioare. Nevoia noastră de a reduce incertitudinea ne guvernează privirea, de exemplu, astfel încât mișcările rapide ale ochiului scanează zonele unei scene în care există o probabilitate mai mare să fie găsite informații mai precise[6]. Mai simplu, semnalele relativ puternice atrag atenția: li se atribuie o precizie mai mare.

Așa funcționează predominanța. Când sunt testate, caracteristicile „predominante” ale lumii minimizează incertitudinea referitoare la ipoteza prioritizată în mod curent a sistemului: ele sunt cele care, atunci când lucrurile se desfășoară după cum s-a preconizat, ne maximizează încrederea în ipoteză. Astfel, agenții activi sunt determinați să eşantioneze lumea ca să (încerce să) își confirme propriile ipoteze[7].

Întrucât acestea sunt, în fond, ipoteze despre cum să ne satisfacem nevoile, asta înseamnă că fiecare specie este determinată să-și selecteze propria lume perceptivă. Orientarea perceptivă a fiecărei specii este dictată de lucruri care contează pentru ea. În consecință, oamenii, rechinii și liliecii trăiesc în lumi (subiective) diferite. Percepi obiecte și evenimente numai atunci când le observi, iar pentru fiecare specie sunt predominante altele. Nu poți să vezi decât ceea ce testezi[8].

Biologul chilian Francisco Varela a exprimat acest lucru foarte bine: „Specia își prezintă și își specifică propriul domeniu de probleme”[9].

Asta presupune că precizia nu poate fi determinată în mod pasiv – nu putem doar să așteptăm ca să vedem ce semnale sunt puternice fără să avem niciun fel de așteptare. Precizia trebuie să fie inferată, apoi *atribuită* de modelul generativ. Atenția – care nu poate fi separată de precizie – poate să fie, în consecință, atât „captată”, cât și „direcționată”[10]. De exemplu, reducem în mod activ la aproape zero precizia în ceea ce priveşte erorile când ne culcăm, dar un eveniment suficient de surprinzător tot ne va trezi. În alte situații, e posibil să o amplificăm în mod activ, ca atunci când ne concentrăm intens asupra unui text dificil, pentru că bănuim că el conține ceva important.

În articolul (2018) care a propus teoria noastră despre conștiință, am schimbat împreună cu Friston unele dintre simbolurile convenționale folosite în ecuațiile ce calculează energia liberă (vezi notele de final ale capitolului 7). Le-am înlocuit ca să precizăm că păşim pe urmele lui Sigmund Freud, care a încercat în 1895 să ofere „o psihologie științifică, adică” să prezinte „procesele psihice ca pe niște stări determinate cantitativ ale unor particule materiale detectabile”, astfel făcând aceste procese clare și lipsite de contradicții[11].

Acestea erau primele rânduri ale „Proiectului de psihologie”. Aici, Freud a folosit simbolurile ϕ, ψ, ω și M pentru a indica patru sisteme ipotetice de neuroni responsabile pentru percepție, memorie, conștiință și, respectiv, acțiune; și a folosit simbolul Q pentru a denota stimulii externi. Eu și Friston am

folosit această notație ca să denotăm vectorii echivalenți în cadrul unui sistem care se autovalidează[12]:

$Q\eta$ = stări externe, așa cum sunt modelate de stările interne ale sistemului

ϕ = stări senzoriale

M = stări active

ψ = predicții

ω = precizii

În plus, noi am folosit:

e = erori de predicție (bazate pe ϕ și pe predicția sa ψ)

F = energie liberă Friston (bazată pe e și pe precizia ω)[13].

De observat că niciuna dintre aceste cantități nu măsoară direct stările externe, deoarece sunt *ascunse* de un sistem cu autoorganizare. Asta înseamnă, în tot ce urmează, că avem o descriere autonomă, independentă a dinamicii psihice în ce privește propriile stări interne ale sistemului ($Q\eta$, ω) și stările zonei Markov (ϕ, M). Fiind acum echipați cu acești termeni, putem să formalizăm dinamica unui sistem ce se autovalidează în relaţie cu optimizarea preciziei.

Voi începe cu două ecuații care definesc energia liberă Friston în termenii cantităților pe care le-am prezentat. Prima ecuație spune că „energia liberă este (aproximativ) logaritmul negativ al probabilității de a întâlni unele stări senzoriale cu autor activ”[14]. A doua spune că „energia liberă preconizată scade (aproximativ) proporțional cu precizia logaritmului negativ”[15]. Amintește-ți că dinamica sistemului care se autovalidează există ca să minimizeze energia liberă.

După ce am stabilit aceste relații, devine evident că de fapt

există *trei* moduri în care un sistem ce se autovalidează reduce eroarea de predicție și, prin aceasta, minimizează energia liberă – nu doar cele două, evidente, pe care le-am descris mai sus:

1) Poate *să acționeze* (adică să schimbe M) ca să modifice senzațiile (ϕ) astfel încât ele să se potrivească cu predicțiile sistemului. Aceasta este acțiunea.
2) Poate *să-și schimbe reprezentarea lumii* ($Q\eta$) ca să genereze o predicție mai bună (ψ). Aceasta este percepția.

Iar acum, în plus:

3) Poate *să ajusteze precizia* (ω) ca să se potrivească în mod optim cu amplitudinea erorilor de predicție primite (e).

Aceasta afirm eu că este conștiința[16].

Acest proces de optimizare finală, optimizarea *încrederii* sistemului, așa cum este descris în textul de mai sus, este cel pe care eu și Friston îl asociem cu evaluarea energiei libere ce stă la baza experienței resimțite. Ecuațiile care formalizează aceste dinamici sunt prezentate în notele de final[17]. Întrucât a treia ecuație este cea esențială, o voi exprima în cuvinte: „În timp, rata schimbării preciziei (ω) depinde de cât de multă energie liberă (F) se schimbă când schimbi precizia. Asta înseamnă că precizia va părea că încearcă să minimizeze energia liberă[18]. Rata acestui proces de minimizare a energiei libere este diferența dintre variație (precizia inversă) și suma erorilor de predicție la pătrat ($e.e$)”[19]. În termeni mai simpli, a treia ecuație cuantifică modul în care *ajustarea continuă a preciziei implementează legea lui Friston*, alături de acțiune și percepție. Pe scurt, cuantifică modul în care *conștiința contribuie la*

acțiune, la percepție și la actualizarea modelului – și, prin aceasta, la minimizarea energiei libere.

Figura 17 este o reprezentare grafică a acestor dinamici.

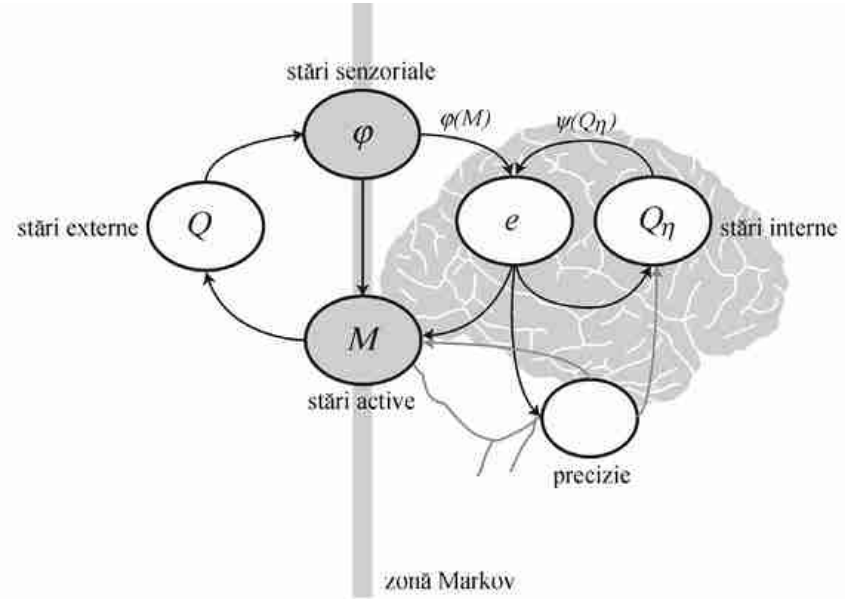


Figura 17. Dinamica unui sistem care se autovalidează, echipat cu optimizarea preciziei. Simbolurile sunt explicate în text. (*Q* este realitatea externă, care este ascunsă de sistem și, prin urmare, nu apare în ecuații.)

La nivel conceptual, precizia este un determinant-cheie al minimizării energiei libere și, prin urmare, al activării erorilor de predicție. Precizia determină ce erori de predicție sunt selectate – și de aceea, în ultimă instanță, modul în care reprezentăm lumea și acțiunile noastre asupra ei. Dacă precizia este „activare” (ceea ce și este), asta explică, din punct de vedere formal și mecanicist, de ce optimizarea încrederii este întotdeauna și exclusiv un proces endogen. Conștiința *nu are cum să vină decât* din interior[20].

Procesul pe care l-am prezentat aici se manifestă în multe moduri. Să-l mai privim o dată în termeni biologici (adică fiziologici și psihologici). În domeniul *exteroceptiv*, se manifestă ca atenție și atenuare, asociate cu creșterea și scăderea preciziei senzoriale[21]. În domeniul *proprioceptiv*, corespunde preciziei potențialităților motoare (utilizări posibile ale obiectelor), de tipul asociat cu selectarea și realizarea obiectivelor[22]. În domeniul *interoceptiv*, acest proces determină concret „senzațiile viscerale”, deci cea mai bună explicație pentru semnalele interoceptive care au fost facilitate sau activate[23]. Totuși, e foarte important să observăm că toate aceste lucruri (exterocepție, propriocepție și interocepție) pot să apară fără conștiință; conștiința este *simțirea* acestor lucruri.

Pe scurt: sarcina preciziei este *să activeze* reprezentări (și așteptări). În lipsa preciziei, erorile de predicție nu vor reuși să inducă nicio sinteză perceptivă sau un comportament motivat. Cu alte cuvinte, fără precizie, erorile de predicție ar fi izolate acolo unde se formează. Asta s-a întâmplat, de exemplu, la pacienții cu achinezie și mutism ai lui Oliver Sacks.

Această formulare a preciziei implică mecanismele neuromodulatoare descrise în capitolul 6, care generează stări alterate de conștiință[24] și vise[25] și care sunt ținta medicamentelor ce modifică conștiința (de exemplu, psihotropele și psihedelicele)[26]. De asemenea, această formulare a preciziei dă o oarecare validitate versiunilor neuromodulatoare ale teoriei conștiinței ca „spațiu de lucru global” (discutată pe scurt în capitolul 4)[27].

Deloc surprinzător, rolul preciziei în psihopatologie este o temă importantă în domeniul în plină dezvoltare al psihiatriei computaționale[28]. Să ne amintim de cazul domnului S., în care

am văzut ce se întâmplă când semnalelor de eroare li se atribuie o pondere prea mică. Putem vedea din figura 17 cum anume triunghiul său de decizie și, prin urmare, sistemul reticulat activator (ω) au atribuit o pondere prea mare modelului său predictiv ($Q\eta$) și prea mică erorilor sale de predicție (e). După cum am văzut, de aici veneau sentimentele lui.

Știu că pare ciudat să vorbim despre conștiință în asemenea termeni mecaniciști. Dar am apelat la ei pentru că descriu legile aflate la baza experienței fenomenologice, și nu experiența fenomenologică în sine. Prin prezentarea acestor legi, încerc să demonstrez că conștiința face parte din natură, că ea nu există în vreun univers paralel, că nu e ceva dincolo de științe.

Acum te rog să treci un Rubicon împreună cu mine.

Acest capitol abordează motivele pentru care și modul cum apare conștiința, nu dintr-o perspectivă biologică, ci din una *formală și mecanicistă*. Mai precis, pune două întrebări: (1) De ce și cum se leagă procesul de prioritizare a nevoilor biologice pe care l-am descris mai sus de *legile* minimizării energiei libere? (2) De ce și cum se întâmplă că acest proces mecanicist *face* unele sisteme cu autoorganizare să fie ceva cum e? Întrebarea de care m-am ocupat în ultimele câteva pagini, din perspectiva fizicii statistice, este cea dintâi. Așa că acum trebuie să explic direct de ce și cum anume dinamicile statistico-mecanice pe care le-am descris generează experiențe resimțite. Cum poate un simplu proces de prelucrare a informațiilor să simtă? Ca să explic acest lucru, trebuie să te rog să faci un salt pe care mulți specialiști în științele naturale ezită să-l facă – iar din această cauză, știința în general și mai ales știința psihicului

au de pierdut. Mă refer la analiza dinamicilor mecanice pe care le-am descris *din punctul de vedere al sistemului*. Te rog să înlocuiești perspectiva obiectivă, la persoana a treia din care am abordat până acum dinamicile în acest capitol cu una la persoana întâi: cu perspectiva *subiectivă* a sistemului care se autovalidează. Te rog să adopți punctul de vedere al sistemului, *să empatizezi* cu el[29].

Acest salt este justificat de două fapte pe care le-am explicat deja. Primul este că experiența resimțită poate fi consemnată numai din perspectiva subiectivă. A nu lua în considerare perspectiva subiectivă înseamnă așadar a exclude din știință cea mai importantă trăsătură a minții. Asta au făcut behavioriștii, pregătind terenul pentru jumătatea de secol în care neuroștiințele nu au reuşit să abordeze conștiința într-un mod semnificativ. Al doilea fapt este: am arătat, în termeni formali și mecaniciști, cum apare subiectivitatea sistemelor care se autovalidează. Prin urmare, asumarea perspectivei subiective a unui sistem care se autovalidează este justificată tocmai de faptul că posedă ipseitate.

Ca să fie clar ce spun: energia liberă și preciziile sale constitutive pot fi cunoscute numai în interiorul unui sistem când el este înțeles în mod subiectiv, din punctul de vedere al sistemului; experiențele nu pot fi observate *ca experiențe* din exterior, în mod obiectiv.

De asemenea, am explicat în termeni cauzali, principiali de ce și cum ipseitatea unor astfel de sisteme este *intențională*. Sistemele cu autoorganizare cu dinamicile pe care le-am descris au un obiectiv și un scop – să supraviețuiască. Asta înseamnă că ele au un sistem de valori, cel aflat la baza tuturor formelor de viață.

Intenționalitatea sistemelor dinamice cu autoorganizare le obligă să pună întrebări despre propriile stări în raport cu perturbațiile entropice care le înconjoară. Asta le face sisteme „care se autovalidează”. Ele *nu au cum să nu* întrebe mereu: „Ce se întâmplă cu energia mea liberă dacă fac asta?”. Mai mult, sistemele complexe ce se autovalidează (ca noi, vertebratele) trebuie să pună această întrebare în relație cu *multe* variabile categoriale; deci răspunsurile – statisticile noastre vitale – trebuie să fie atât cuantificate, *cât și calificate*. În sfârșit, ele trebuie să-și moduleze nivelul de *încredere* în răspunsurile pe care le primesc.

Ceea ce descriu aici în termeni abstracți, tehnici nu este prea complicat. O știi din experiență personală. Ceea ce trăiești tot timpul sunt impulsuri fluctuante a ceea ce simţi ca răspuns la mișcarea ta prin lume, pe măsură ce verifici dacă totul e așa cum te așteptai să fie – și pe măsură ce încerci să reduci cumva decalajul atunci când nu e cum te așteptai. Practic, nu asta înseamnă experiență pentru tine?

Prin combinarea tuturor acestor fapte despre subiectivitatea și intenționalitatea sistemelor complexe care se autovalidează, ajungem la concluzia următoare. Răspunsurile evocate de echipament (în sensul lui Wheeler) care circulă în mod subiectiv pornind de la tipurile de întrebări la care sistemele ca noi sunt obligate să răspundă trebuie să aibă valoare existențială și mai multe calități. Încrederea noastră în aceste răspunsuri fluctuante – „fenomenele” – pe care le consemnăm trebuie să fie subiectivă, calificată și să aibă valențe.

Și asta *înseamnă pur și simplu* ceea ce înseamnă să experimentezi în mod conștient. Răspunsurile evocate de echipament, cel puțin în cazul vertebratelor ca noi (și, fără

îndoială, și al altor organisme), sunt simțite.

Ca să te ajute să treci Rubiconul, amintește-ți, te rog, că sentimentele *au evoluat*. Zorii conștiinței au dat naștere unor fenomene foarte simple, precum senzația că îți este prea cald. Dovada senzorială tot mai precisă care prezice moartea unui sistem ce se autovalidează care se supraîncălzește pur și simplu este resimțită de sistem ca „o căldură prea mare”. A trecut o veşnicie până s-au dezvoltat asemenea forme elementare de afect, în cadrul unei ierarhii predictive profunde, creând, în cele din urmă, „o lume tridimensională, panoramică, bine definită, compusă din obiecte solide cu formă clară: lumea experienței noastre fenomenologice familiare”, ca în viziunea lui Merker.

Într-o asemenea lume, simțim ce înseamnă să fii un sistem cu dinamicile pe care le-am descris. Sentimentele sunt stări subiective fluctuante, cu valoare existențială, cu calități diferențiate și cu grade de încredere. Aceasta este materia conștiinței. Acum înțelegem de ce nu are cum să fie altfel.

10

Înapoi la cortex

După cum am văzut în mod repetat în călătoria noastră, eroarea privitoare la cortex are de dat socoteală pentru multe. Dacă pionierii neuroștiințelor behavioriste nu ar fi fost atât de impresionați de marea întindere a cortexului nostru sau dacă nu ar fi fost atât de orbiți de ideea filosofică potrivit căreia viața psihică apare din asocierea unor imagini mnezice, poate că am fi descoperit de mult adevărata sursă a conștiinței. În istoria științei psihicului, e o ironie ce nu ne dă pace că Freud avea atât de multe piese ale puzzle-ului acum mai bine de un secol. Avea sub nas indiciile, atât neurologice, cât și psihologice. Dar, când a fost vorba de conștiință, până și el a căzut pradă fixației noastre colective pentru cortexul cerebral – obsesie al cărei preţ, în caz că uităm, poate fi măsurat nu doar în timp pierdut.

Toate acestea sunt adevărate. Totuși, cortexul are în mod clar un rol masiv. Experiența noastră de zi cu zi este strâns legată de dinamica procesării corticale. Așa că, în acest capitol, să ne întoarcem la etajul superior înțeles greșit al creierului, ca să vedem ce adaugă el la descrierea pe care o facem aici conștiinței. Vom afla că multe dintre trăsăturile cele mai banale ale experienței noastre de zi cu zi își derivă caracterul din ceea

ce face cortexul – dar nu în modurile la care ne-am gândit până acum.

Sensul cel mai evident în care este valabilă afirmația de mai sus e că lumea așa cum o resimțim se generează concret din reprezentări corticale. Deși pare ciudat, în cadrul codificării predictive ceea ce percepem este o *realitate virtuală* clădită din propriile materiale de construcție ale minții.

Această afirmație este radicală în comparație cu perspectiva simțului comun, dar ideea că experiența perceptivă este autogenerată se bucură de o largă recunoaștere în neuroștiințele contemporane. De exemplu, să ne amintim ce a spus Semir Zeki despre vederea în culori încă din 1993. Zeki a scris că culoarea este „*o proprietate a creierului*, o proprietate cu care el învestește suprafețele exterioare, o interpretare pe care o dă anumitor proprietăți fizice ale obiectelor”[1]. Apoi a elaborat:

> Să presupunem că privim o mică suprafaţă scăldată izolat de o lumină cu unde lungi [...]. Suprafaţa produce un spectru larg de luminozitate pentru lumina cu orice lungime de undă, întrucât singura comparație pe care o poate face creierul în aceste condiții este între lumina ce se reflectă din zona iluminată și cadrul întunecat din jur. Astfel, lumina cu unde lungi produce luminozitate mare, iar cea cu unde medii și scurte, fiind absentă, nu produce luminozitate deloc. Astfel, sistemul nervos atribuie zonei culoarea roșie[2].

De observat cum își alege Zeki cuvintele: creierul *atribuie* lumii culoarea roșie, după ce a pus întrebări despre intensitățile relative și lungimile de undă ale luminii. Face o

pictură a lumii pe numere. Același lucru se aplică şi proprietăților fenomenologice care ne caracterizează celelalte modalități de percepție: sunetele, gusturile, senzațiile somatice și mirosurile. Creierul atribuie lumii aceste calități.

Presupun că multor cititori le vine greu să creadă că ceea ce vedem chiar în acest moment nu este pur și simplu ce e „acolo”. Îmi imaginez că întrebi: „De unde altundeva îmi vine perceperea cuvintelor de pe pagina asta?”. Poate te ajută dacă precizez că ce vezi acum nu prea seamănă cu inputurile senzoriale pe care le primești. La început, aceste inputuri sunt unde de lumină ce ajung pe retină. Aici, celulele fotosensibile (numite conuri și bastonașe) răspund la undele de lumină generând impulsuri nervoase. Aceste impulsuri – și nu undele de lumină – sunt apoi propagate de-a lungul nervilor optici spre cortex sub forma unor *spike trains* (vezi figura 11). De ce resimți aceste *spike trains* – 001111101101 – ca imagini mișcătoare aflate în lume?

Neuronii din corpul geniculat lateral și din cortexul „de proiecție” occipital ce răspund la impulsurile retiniene au o aranjare topologică, creându-se astfel posibilitatea formării unor imagini prin *cartografierea* suprafețelor retiniene (vezi figura 6). Acest lucru însă nu ține cont de faptul că nu există bastonașe și conuri în apropiere de centrul retinei, de unde porneşte nervul optic. Atunci, s-ar justifica să vezi o gaură neagră mai spre centrul câmpului vizual. Cum dispare această gaură? Răspunsul e că *inferezi* din context și din memorie ce este în „pata oarbă”, apoi o umpli[3].

Ar fi trebuit să spun că există o gaură neagră în câmpurile tale vizuale (la plural) pentru că, să nu uităm, ai două. De aici apare altă întrebare: de ce nu vezi două imagini? Nu mă refer

la faptul că ai doi ochi. E simplu să ne imaginăm cum se pot suprapune două imagini aproape identice, dar nu asta se întâmplă[4], ci celulele din jumătatea stângă a fiecărei retine proiectează în lobul occipital drept, iar cele din jumătatea dreaptă proiectează în lobul stâng. Asta înseamnă că în cortexul vizual ai de fapt două reprezentări diferite ale suprafețelor retiniene (una a jumătății din stânga a acestei cărți și una a jumătății din dreapta)[5], cu o fisură anatomică între ele: fisura longitudinală care împarte emisferele cerebrale. Cum anume devin cele două câmpuri imaginea unitară pe care o vezi? (E adevărat că sunt coordonate de axoni în corpul calos, dar și oamenii cărora le lipsește corpul calos văd o singură imagine[6].) În plus, trebuie să ținem cont de faptul că câmpurile vizuale – așa cum sunt ele reprezentate în lobii occipitali – sunt cu spatele în față și cu susul în jos față de imaginile pe care le vezi. Mai mult, ochii tăi fug în toate direcțiile, cam de trei ori pe secundă, ca să nu mai vorbim de mișcarea constantă a capului. Cum se face că percepi o scenă vizuală stabilă și orientată corect?

Ideea mea – că există puține asemănări între ceea ce vezi și inputurile senzoriale ce ajung în cortex – este viu ilustrată de pacienții cu antecedente neurologice cărora le sunt afectate mecanismele prin care transformăm în mod normal ceea ce ajunge acolo în ceea ce vedem. Eu am prezentat un asemenea caz cu mulți ani în urmă: un băiat de 12 ani (W.B.) cu abcese bilaterale la lobii frontali, cauzate de pansinuzită[7]. Periodic, pacientul vedea lumea rotită la 180°. Simptomele și semnele acestui pacient erau identice cu cele descrise în 21 de rapoarte ale unor cazuri anterioare pe care le-am găsit prin literatura de specialitate începând din 1805, conferind credibilitate descrierii

lui subiective (vezi figura 18 pentru unele dovezi obiective).

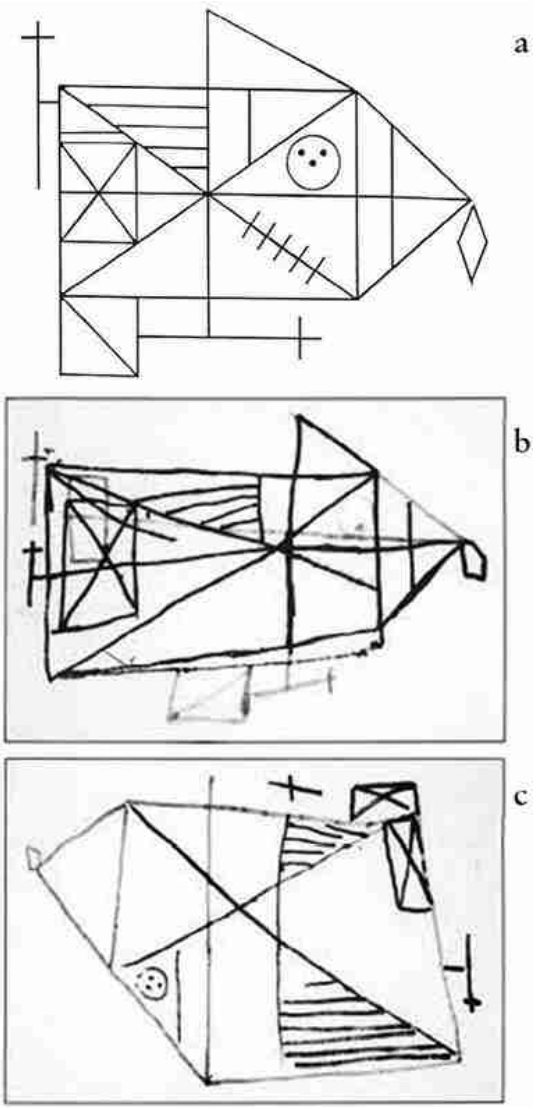


Figura 18. (a) = figura complexă Rey, așa cum i-a fost arătată pacientului W.B.; (b) = aceeași figură copiată de pacient; (c) = figura desenată din memorie de pacient. Aceste desene constituie o dovadă obiectivă a inversiunii modelului predictiv al lumii pe care îl avea W.B. Figura complexă din testul Rey este dificil de reprodus chiar și în cele mai bune momente; de ce i-ar crește dificultatea acest pacient foarte bolnav, desenând-o răsturnată?

În prezent, mă ocup de o pacientă și mai interesantă, care este inclusă într-un studiu de Aimee Dollman, o doctorandă a mea. Nu pot să prezint toate detaliile, deoarece cazul este încă în studiu, iar concluziile nu sunt publicate. Pacienta este o tânără foarte inteligentă cu disgenezie cerebrală (anatomie anormală) a lobilor occipitali, ce îşi reprezintă lumea aproape așa cum tocmai am spus că *nu* o trăim: așa cum este aranjată din punct de vedere anatomic în cortexul vizual. Asta se întâmplă mai ales când își folosește memoria vizuală (adică modelul ei predictiv). Vede două câmpuri separate, cu spatele în față și cu susul în jos (nu întotdeauna la unison unul cu celălalt). Acest model vizual al lumii *nu reușește să facă inferențele corectoare obișnuite* prin care orientăm și integrăm câmpurile vizuale. Prin urmare, acesta face predicții eronate, iar experiența ei vizuală nu se potrivește cu cea a celorlalte modalități senzoriale. În consecință, uneori devine confuză în ce privește direcția în care se mișcă corpul ei (mai ales când călătorește cu trenul sau cu avionul) și face erori majore când se orientează prin mediul ambiant. În plus, îi vine greu să infereze care sunt obiectele constante și care sunt datele vizuale pasagere (cum ar fi ortografierea unor cuvinte în minte când trebuie să le abstragă din fragmente și fonturi variabile sau identitatea unor chipuri abstrase în condiții de lumină schimbătoare și din perspective vizuale diferite).

Cu alte cuvinte, cortexul ei vizual primește tipul de informații pe care le primim cu toții, dar ea nu poate să generalizeze pornind de la semnalele senzoriale sonore și nici să infereze automat obiectele stabile reprezentate de acestea (de exemplu, să recunoască un chip familiar). Ea a suferit toată viața din cauza acestor anomalii și și-a dezvoltat metode

elaborate prin care să le compenseze. Întrucât cortexul ei de „asociere” vizuală nu integrează *automat* cele două câmpuri vizuale, rotind scena, pacienta își ajustează reprezentarea lumii făcând inferențe *deliberate*. De exemplu, când am rugat-o să identifice locația unui oraș bine-cunoscut pe o hartă nemarcată, a spus: „Să vă arăt unde *simt* că este sau unde *știu* că e?”. Atunci când „simte” că ceva se află la vest, ea „știe” că se află probabil la est.

Totuși, în loc să apelez la asemenea afecțiuni neuropsihologice rare, mai bine să ilustrez natura autogenerată a percepției cu ajutorul fenomenului utilizat de regulă în acest scop, şi anume *rivalitatea binoculară.*

Acest fenomen a fost descris pentru prima dată în 1593 și a ocupat un loc important în opera fundamentală a lui Helmholtz pe tema inferenței inconștiente[8]. Fenomenul implică prezentarea simultană a unor imagini diferite fiecărui ochi, cu ajutorul unui stereoscop cu oglinzi. Să spunem că ochiului stâng îi este prezentat un chip, iar celui drept o casă. În aceste condiții artificiale, experiența vizuală se desfășoară într-un mod „bistabil”, în care nu vezi o suprapunere a celor două imagini, ci o *alternare* între ele. Vezi o casă, apoi un chip, apoi o casă, apoi un chip, și nu o imagine combinată casă-chip. Asta demonstrează clar deosebirea dintre semnalul obiectiv transmis creierului și percepția subiectivă generată de el. Helmholtz a concluzionat: „În asemenea cazuri, *interpretarea* [semnalului vizual] oscilează astfel încât observatorul are experiențe diferite, una după alta, pentru imaginea retiniană ce nu se schimbă”[9]. Prin urmare, ca și în cazul vederii în culori, experimentezi o *inferență* despre inputul senzorial, și nu inputul în sine.

În viața de zi cu zi e în general cam la fel, ca atunci când „am văzut-o” pe prietena mea britanică Teresa la aeroportul din Cape Town. Toate aceste iluzii arată că ceea ce percepi este generat în mare parte de *așteptările* tale. În termeni bayesieni, se consideră că rivalitatea binoculară arată că, dacă *ipoteza anterioară* ce se potrivește cel mai bine cu datele senzoriale (verosimilitatea mare că vezi o casă-chip) nu corespunde *cunoașterii de fond* (probabilitate mică să existe case-chipuri), atunci ipoteza este respinsă. Inferența că vezi o casă domină inferenţa că vezi o casă-chip, așa încât ceea ce vezi este o casă. Dar, atunci când testezi această ipoteză posterioară (ca o nouă ipoteză anterioară), ea se potrivește numai cu jumătate dintre dovezile senzoriale. Cunoașterea ta de fond îți spune că chipurile întregi sunt la fel de probabile precum casele întregi. Așa că te răzgândești și inferezi că probabil vezi un chip – apoi asta vezi. Însă, atunci când testezi această ipoteză nouă, ea nu explică, din nou, decât jumătate dintre dovezile senzoriale. Și așa mai departe…

Interpretarea bayesiană a rivalității binoculare este larg acceptată. Pe mine mă convinge pentru că, atunci când ni se arată două imagini care, în forma combinată, au o probabilitate anterioară mare, percepem o imagine amestecată. De exemplu, dacă ți se prezintă două imagini astfel încât un ochi vede un canar, iar celălalt o colivie, le vei vedea imediat drept un canar într-o colivie.

Ceea ce percepi nu este totuna cu inputul care vine de la simțuri. *Ceea ce percepi este o inferență*. Iar materialele din care este derivată inferența sunt în cea mai mare parte modelul tău cortical predictiv derivat din experiențe trecute (adică preconizate)[10].

Asta ne spune ceva despre ce aduce cortexul în planul conștiinței. Dar ce putem spune despre situația inversă? Ce face conștiința pentru cortex?

Ceea ce urmează să spun este evident, dar se pare că nu o mai exprimă nimeni[11]: conștiința cognitivă este generată de un mecanism neuronal recent descoperit numit „reconsolidarea memoriei”. La fel ca multe altele, această idee își are originea în scrierile lui Freud, conform căruia „*conștiința ia naștere acolo unde se sfârșește urma mnezică*”[*1] [12]. Ceea ce avea în vedere Freud era ușor diferit de ce urmează să spun eu, deoarece el – ca toți neurologii vremii – era blocat în eroarea privitoare la cortex[13]. Totuşi, şi în acest caz e uimitor cât de aproape de adevăr a ajuns.

Am văzut că afectele înaintează cereri minții, iar cogniția face travaliul astfel solicitat. Mai exact, cogniția *conștientă* face treaba; căci, odată ce sarcina a fost îndeplinită, iar încrederea în credința (prioritizată) ce devenise nesigură a fost restabilită, modelul generativ își reia modul automat de funcționare, sub pragul conștiinței[14]. Aici avem din nou mecanismul *învățării din experiență* pe care l-am tot descris. Acesta este rostul conștiinței în cogniție. Ajungi într-o situație în care nu ești sigur ce să faci. Conștiința vine să te salveze: îți cauți ieșirea din scenariul respectiv în funcţie de ceea ce simţi, observând acțiunile voluntare ce dau rezultate bune în cazul tău. Apoi, treptat, lecțiile de succes devin automatizate, iar conștiința nu mai este necesară[15].

Vreau să evidențiez că travaliul cognitiv pe care tocmai l-am descris *încetinește* activarea în lume, care de altfel e ceva automat. Aceasta este diferența esențială dintre acțiunea

voluntară și cea involuntară, dintre cogniția conștientă şi cea inconștientă, dintre pulsiunea resimțită şi reflexul autonom – ceea ce are caracter voluntar este mai puțin sigur și, de aceea, necesită mai mult timp. Acest proces, care întârzie tendințele de acțiune automatizată și le permite să fie ținute minte (în memoria pe termen scurt), se numește în mod inspirat „memorie de lucru”, care înseamnă, concret, că păstrezi în minte ceea ce simţi – afect stabilizat transformat în travaliu cognitiv. După cum tocmai am spus: dacă afectul înseamnă solicitare de travaliu înaintată minții, atunci cogniția conștientă este travaliul în sine. Astfel, afectul însoțește și totodată *devine* cogniția. „Travaliul” despre care e vorba presupune inhibarea tendințelor de acțiune automată și *stabilizarea* intenționalității în timp ce sistemul își caută ieșirea din problemele neprevăzute în funcţie de ceea ce simte. De aici vin avantaje adaptive considerabile, întrucât sunt facilitate soluții viabile la numeroasele probleme din lumea reală pe care modelele noastre generative (încă) nu pot să le prezică. Acest proces stabilizator este funcția cortexului[16]. Cortexul se specializează în incertitudine.

Ceea ce implică toate acestea este că *starea conștientă este indezirabilă* din punctul de vedere al unui sistem cu autoorganizare. Să revenim asupra diagramei de tip cadran din figura 12: săgeata din exterior reprezintă cererea tot mai mare de travaliu (afect negativ), iar cea din interior cererea în scădere (afect pozitiv), dar starea ideală este punctul de stabilizare ce reprezintă *zero cereri*. În capitolele 7-9, am oferit o fundamentare mecanicistă formală pentru aceste chestiuni. Minimizarea energiei libere este starea ideală a sistemelor vii, însemnând că surpriza minimă este idealul. Adică *nevoia*

minimă este idealul. Afectul nu este decât anunțarea unei nevoi predominante. Asta ar trebui să însemne că sentimentul este ceva bun, deoarece permite sistemelor biologice să-și rezolve nevoile și, prin aceasta, să evite distrugerea. Dar starea *ideală* este cu siguranță una în care toate nevoile noastre sunt satisfăcute automat – chiar înainte să fie simțite –, adică o stare *fără incertitudine*. În această stare ideală teoretică, în care nevoile ne sunt satisfăcute automat, nu simțim nimic. (Așa sunt satisfăcute majoritatea nevoilor noastre corporale: sunt reglate în mod autonom.) Spun „ideală teoretică" deoarece, în ce privește multe dintre nevoile noastre, în special cele emoționale, nu o atingem niciodată. Numai pulsiunea CĂUTĂRII rezolvă acest lucru. Iar asta este vestea proastă.

Vestea bună este că semnalele de eroare cu cele mai mari valori de precizie au influența cea mai mare asupra modelului generativ. Întrucât cer schimbarea cea mai amplă – declară că faci ceva *greșit* –, ele constituie cele mai mari ocazii de a învăța, astfel încât să nu fii la fel de încurcat data viitoare. E același lucru cu „a învăța din greșeli", după cum ți-au spus să faci mereu părinții și profesorii. Acum știi de ce: au vrut să te scutească de uitare. Dar efectul surprinzător rămâne că, în cogniție, *conștiința este indezirabilă*. Prin urmare, toți aspirăm nu la plăcere (nevoia în scădere), ci la starea de zombi (zero nevoi). Zero nevoi implică predicții perfecte, adică zero erori și, de aceea, zero cereri de creștere a preciziei pentru semnalele de intrare și, prin urmare, nimic de simţit. În sfârșit, liniște.

Acum ar trebui să înțelegem mai bine aforismul lui Freud „Conștiința ia naștere acolo unde se sfârșește urma mnezică". Cu alte cuvinte, *conștiința apare atunci când comportamentul automat duce la eroare* – când urma mnezică (o predicție) care

produce un comportament nu are efectul preconizat. Asta înseamnă că predicția respectivă trebuie *actualizată* astfel încât să înglobeze eroarea. Așadar, conștiința corticală poate fi descrisă drept „*activitate predictivă în desfășurare*”. O urmă mnezică ce este conștientă e în curs de actualizare. Nu mai este urmă mnezică. De aici apare conștiință în locul unei urme mnezice[17].

În ultimii ani ai secolului XX, am înțeles neurofiziologia acestui proces în termeni de „eroare de predicție a recompensei” (vezi figura 19)[18]. Începând din primii ani ai secolului XXI, înțelegem mai bine fiziologia actualizării memoriei, numind-o „reconsolidare”.

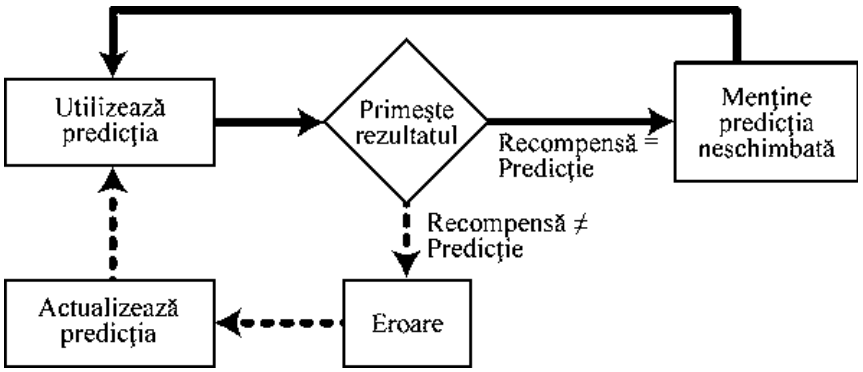


Figura 19. Schemă a învățării prin eroarea de predicție a recompensei. Secvența începe cu caseta „Utilizează predicția”, care duce la două rezultate posibile, în partea dreaptă a casetei. Fie există o eroare de predicție (când rezultatul diferă de predicție), ceea ce duce la caseta „Actualizează predicția”, fie nu există nicio eroare (când rezultatul se potrivește cu predicția), ceea ce duce la caseta „Menține predicția neschimbată”, apoi comportamentul rămâne neschimbat. Etapa „Actualizează predicția” corespunde reconsolidării.

Elementele fundamentale au ieșit pentru prima dată la lumină odată cu descoperirea, în anii 1960, a faptului că amintirea unei frici poate fi eliminată cu ajutorul terapiei electroconvulsive (ECT) – dar numai dacă electroșocul este administrat imediat după accesarea amintirii[19]. Asta sugera că ECT interferează cu un proces ce trimite în starea lor latentă amintirile manifeste (active) ale fricii: dacă electroșocul este administrat în timp ce amintirea fricii este activată, atunci această amintire este ștearsă. Concret, ea *nu mai este o urmă* în timp ce o rememorezi. Starea activată a amintirii face urma pe termen lung să fie *instabilă* din nou: o face să nu mai fie o amintire.

Acest lucru a fost confirmat în anul 2000, când reconsolidarea a fost redescoperită și numită astfel[20]. Dacă inhibitorii sintezei proteinelor sunt administrați în timp ce este activată o urmă pe termen lung, ea dispare. (Inhibitorii sintezei proteinelor împiedică formarea unor *noi* urme pe termen lung.) Asta e valabil și în cazul altor tipuri de memorie, nu numai al amintirii fricii. În general, amintirile pe termen lung devin instabile când sunt în stare activată. În acest fel se actualizează (apoi se consolidează din nou – adică se reconsolidează).

În ultimele două decenii s-au realizat numeroase studii privind reconsolidarea, arătându-se că ea apare nu numai la oameni și la rozătoare, ci și la găini, pești, crabi, melci de apă dulce și albine. Aceste studii arată și existența a ceva analog reconsolidării în căile de procesare a durerii din măduva spinării, sugerând că are un rol primar în sistemul nervos central.

Memoria pe termen lung, spre deosebire de cea pe termen

scurt, depinde de sinteza unor noi proteine, ce se declanșează după transmisia sinaptică substanțială și repetitivă, care, la rândul ei, este modulată de sistemul reticulat activator – adică de activare[21]. De unde și celebrul adagiu numit legea lui Hebb: „Neuronii care se descarcă împreună se și conectează”[22]. O amintire „activată” nu mai este o amintire – se află într-o stare de incertitudine. Tot ce încerc să transmit este că conștiința cognitivă se reduce la inducerea instabilității la nivelul urmelor memoriei corticale și că această instabilitate este un *produs al activării*. Tot ajungem din direcții diferite la aceeași concluzie: procesele corticale sunt chestiuni fundamental inconștiente (sunt pur și simplu niște algoritmi, dacă sunt lăsate în voia lor). Conștiința – toată – vine din trunchiul cerebral[23].

Cu toate acestea, pare că ne străduim să atingem un fel de stare de zombi. Forma ideală de cogniție este automatismul și, cu cât reușim să scăpăm mai repede de conștiință, cu atât mai bine. Și atunci cum devine *inconștientă* cogniția corticală?

Să ne gândim la cel mai simplu exemplu posibil. Ce crezi că s-ar întâmpla dacă aș proiecta o linie verticală în fața ochilor tăi astfel încât să nu iau în calcul mișcările constante ale ochilor, care, în mod normal, garantează că tot ce privești este actualizat cam de trei ori pe secundă? Această actualizare se petrece la periferia ierarhiei predictive, unde se găsesc cele mai mari grade de libertate. Ritmul frenetic în care se actualizează modelul necesar în asemenea circumstanțe impredictibile face cuvântul „predicție” să-şi piardă sensul aproape în întregime. Iată de ce atât de mult travaliu cognitiv se desfășoară la periferia senzorimotoare.

Așadar, ce s-ar întâmpla dacă am putea să imobilizăm un

stimul vizual în așa fel încât să nu-l vedem decât pe el şi asta într-un mod cu totul monoton? Răspunsul tulburător este că *el dispare din conștiință*. Chiar dacă stimulul e tot acolo, el dispare din conștiința vizuală în câteva secunde. Acest lucru a fost demonstrat în anii 1950 de Lorrin Riggs în America și de Robert William Ditchburn în Anglia. De atunci, au fost observate efecte similare și la alte modalități senzoriale[24]. De-acum ar trebui să fie evident motivul pentru care stimulul dispare treptat din conștiință: el devine 100% predictibil, deci nu mai este purtător de informații. Predicția atinge precizie totală, iar eroarea aferentă se reduce la zero. După cum am spus, acesta este idealul homeostatic al creierului.

E dificil să stabilizezi niște stimuli în raport cu mișcările ochilor și capului, așa că nu poți face acasă experimentul Riggs-Ditchburn. Dar poți să cauți pe Google iluzia „lilac chaser”, care prezintă alt tip de estompare vizuală[25]. După ce ai dat clic pe imagine, menține-ți privirea cam la 20 de centimetri de ecranul calculatorului și concentrează-te asupra țintei din mijlocul cercurilor care se rotesc. Apoi observă ce se întâmplă cu cercurile. Ele dispar – deoarece creierul vizual inferează că sunt zgomot și le atribuie mai puțină precizie în raport cu fondul gri, făcându-le să dispară treptat.

Pe site găsești încă două iluzii optice. Pe lângă estompare, vei observa un cerc liliachiu care se transformă într-unul verde ce nu este acolo de fapt – (1) nici nu este cu adevărat prezent și (2) nici nu e chiar verde – și care se învârte în jurul țintei centrale. Pe site se explică ce mecanisme funcționează. Practic, e vorba de cântărirea preciziei. Toate aceste fenomene susțin ideea că ceea ce vezi e generat de creierul tău, nefiind primit din exterior[26].

Să identificăm legătura dintre cele de mai sus și secțiunea din capitolul 5 despre actualizarea elementelor motoare anterioare înnăscute: reflexele și instinctele. Aceste predicții înnăscute ne ajută, dar ele nu au cum să descrie corect toate complexitățile lumii, așa că trebuie completate prin învățarea din experiență. Învățarea necesită conștiință, întrucât ne crește treptat încrederea în predicțiile noastre nou-dobândite. Dar idealul oricărei învățări este *să automatizăm* și aceste predicții dobândite, să le facem să se comporte precum reflexele și instinctele. Sperăm să ne formăm predicții noi care sunt cel puțin la fel de sigure și de generalizabile ca şi cele vechi. În măsura în care putem realiza acest lucru, predicțiile dobândite se automatizează prin *consolidarea* sistemelor. În acest sens, consolidarea este opusul reconsolidării, care anulează urmele consolidate, dizolvând la propriu proteinele care le-au „conectat”.

Asta se întâmplă până la nivelul sistemelor memoriei nondeclarative. Obiectivul învățării din experiență este trecerea cât mai multor amintiri pe termen lung din starea declarativă în cea nondeclarativă, pentru că „declarativ” înseamnă „capabil să revină în conștiință”. Așadar, când am spus că starea de zombi este idealul cogniției, am vrut să spun că idealul învățării este consolidarea tot mai profundă. Memoria nondeclarativă este *cea mai sigură* formă de memorie. Ea presupune cel mai mic volum de travaliu. Minimizează complexitatea și este cea mai generalizabilă (vezi capitolul 8). În consecință, e cel mai rapid de executat; implică cel mai puțin incertitudinea și, de aceea, întârzierea cea mai mică.

Desigur, există aici unele complexități. În primul rând, nu vreau să dau impresia că întreaga consolidare pornește de la

sistemele de memorie declarativă la cele de memorie nondeclarativă. Multe predicții pe termen lung sunt consolidate *direct* în memoria nondeclarativă și majoritatea tipurilor de învățare au loc simultan în ambele sisteme. În al doilea rând, există multe tipuri de memorie nondeclarativă și nu toate funcționează la fel. De exemplu, învățarea „procedurală” are loc prin repetiție – prin urmare, spunem că abilitățile și obiceiurile de tipul mersului pe bicicletă „se învață greu și se uită greu”. Totuși, unele tipuri de răspunsuri emoționale nondeclarative, ce sunt la fel de greu de uitat, se dobândesc prin învățare printr-o singură expunere: condiționarea fricii, de exemplu. (Alte tipuri de învățare emoțională sunt mai lente; atașamentul afectiv, de pildă, se face cam în șase luni.) „A fi greu de uitat” e trăsătura cardinală a memoriei nondeclarative, dar consolidarea presupune procese foarte diferite în numeroasele sisteme ale memoriei. În sfârșit, amintirile nondeclarative sunt „inconștiente” numai în sensul *cognitiv*. Atunci când este declanșat un răspuns emoțional dobândit, simți ceva, doar că nu știi despre ce este vorba – adică de unde a venit (vezi capitolul 3).

Cel mai important fapt despre memoria nondeclarativă e că este nondeclarativă. Ea generează *răspunsuri* procedurale, în timp ce memoria declarativă generează *imagini* experiate[27]. Acest lucru coincide cu o deosebire anatomică: amintirile declarative sunt corticale, iar cele nondeclarative sunt subcorticale[28]. Urmele mnezice subcorticale nu pot fi recuperate ca imagini pentru că nu constau din reprezentări corticale ale organelor senzorimotoare terminale. Ele presupun stereotipuri mai simple, de felul celor pe care le-am descris în legătură cu comportamentele învățate ale copiilor cu

hidroanencefalie și ale animalelor supuse procedurii de decorticare cerebrală[29]. Asemenea lucruri nu pot fi aduse în conștiință, ele „nu pot fi gândite”.

În schimb, sistemele memoriei corticale sunt întotdeauna gata să ruleze scenariile predictive pe care le reprezintă – concret, să le retrăiască. Altfel spus, memoria declarativă livrează prompt urme pe termen lung stării pe termen scurt a memoriei de lucru conștiente. Face asta nu numai ca să actualizeze predicțiile corticale, ci și ca să ghideze acțiunea în condiții de incertitudine.

Urmele mnezice subcorticale sunt mai sigure decât cele corticale – în cazul valorilor lor mari de precizie, probabilitatea să se schimbe este mai mică – deoarece sunt optimizate pentru simplitate, și nu pentru acuratețe. Asta le face mai generalizabile. Dar există un preț: modelele mai puțin complexe sunt mai puțin precise când contextul variază[30]. Complexitatea relativă a predicțiilor corticale, în schimb, coincide cu mai multă plasticitate. Pe scurt, cortexul se specializează în *contexte* – reface acuratețea modelului în situații imprevizibile[31]. Un schimb este inevitabil. Cu cât e mai mare potențialul de experiență conștientă, cu atât mai redus e automatismul, ceea ce înseamnă mai multă plasticitate, dar și mai mult travaliu cognitiv. Asta consumă energie și generează sentimente, așa încât creierul face minimul necesar. Chiar și dacă asta înseamnă să estompeze până la anulare un stimul aflat sub ochii tăi.

Și totuși, o mare parte din ce se întâmplă în capul nostru pare greu de împăcat cu acest ideal de eficiență informațională și termodinamică. Pe lângă sentimente și percepții, conștiința

noastră conține, evident, *gânduri*. Ele au origine corticală, bineînțeles – dar ce sunt? Și de ce par adesea atât de inerte?

Teoria cogniției pe care o prezint aici se învârte în jurul capacității sistemelor noastre mnezice de a genera o lume virtuală[32]. Fiecare nivel din ierarhia predictivă – inclusiv fiecare nivel al procesării corticale – este capabil să genereze o versiune a datelor pe care se așteaptă să le primească de la nivelul de după el. Asta înseamnă că percepția nu e fundamental diferită de imaginație: din punct de vedere subiectiv, nu prea există diferențe între lumile pe care le trăiești în vise și cea pe care o vezi la geam[33]. Creierul poate să evoce la comandă realități fantasmagorice. Probabil o face chiar acum, în timp ce citești aceste cuvinte și te gândești la ce-or fi însemnând. Chiar în timp ce îți lași mintea să hoinărească…

Faptul că mintea poate să hoinărească ar părea o provocare pentru teoria conștiinței din perspectiva energiei libere. Am spus că conștientizăm numai semnalele de eroare prioritizate, puternice – cele *predominante* –, la care trebuie să răspundem ca să ne putem păstra parametrii biologici în limitele viabile. Cu toate acestea, gândurile noastre par adesea să fie aleatorii și nesemnificative. În unele cazuri, ni se pare chiar că monologul nostru interior ne deranjează sau ne distrage: nu ne ajută în situația în care ne aflăm. Cum poate așa ceva să minimizeze energia liberă?

Deși poate părea ciudat, scopul hoinărelii minții este să ne îmbunătățească *eficiența* modelului generativ. După cum postulează principiul energiei libere, un model este eficient numai dacă folosește minimul de resurse necesare pentru a se autoorganiza. Asta se reduce la a găsi modelul cel mai simplu

care prezice cu succes mostrele senzoriale din lume. (Să ne amintim de briciul lui Occam.)

Modelul cel mai simplu nu apare în mod natural dintr-o acțiune voluntară. Acțiunea voluntară este un proces aleatoriu. Prin urmare, simplitatea crește prin „*toaletarea*” conexiunilor sinaptice redundante ce s-au format când învățam din experiență. Le scurtăm ca să evităm „supraajustarea” modelelor noastre la date sonore, prin care am păstra inutil corelații excentrice și slabe. Instrumentele cu care facem toaletarea sunt mecanismele de-acum familiare ale consolidării și reconsolidării memoriei: prin activarea amintirilor, putem să le întărim, să le modificăm și chiar să le ștergem.

Hoinăreala minții este un mijloc prin care se realizează acest lucru. Implică o activitate spontană a prozencefalului (numită și „stare de repaus” sau „mod implicit”), ce apare în lipsa oricărui stimul extern specific. Acest tip de activitate continuă în cea mai mare parte a timpului, în fundal, printr-o „explorare imaginativă a propriului spațiu psihic”[34]. Între vis și această formă de gândire există o suprapunere considerabilă[35], ce pare să survină la toate creaturile dotate cu cortex: orice animal cu capacitatea de a genera imagini ale sinelui acționând în lume poate de asemenea să hoinărească prin nesfârșite lumi simulate, în funcție de cum permit circumstanțele[36]. Poate îți amintești că hoinăreala este strâns legată de pulsiunea CĂUTĂRII, ce își continuă cererile în timp ce dormim. Ar trebui să fie evident de ce activitatea în modul implicit este mai lipsită de riscuri noaptea, când nu e necesar să ne confruntăm cu evenimente exterioare.

Toate acestea explică faptul aparte că modelele lumii pe care le dobândim nu sunt niciodată stabile – nici chiar în timp ce

dormim. Chiar și în lipsa unor inputuri senzoriale convingătoare, activitatea neuronală structurată continuă, generând cercetarea și testarea neîntrerupte ale modelului generativ. Andy Clark tinde să creadă că aceste explorări ar putea duce la noi răspunsuri elegante la problemele care ne ocupă atenția în starea de veghe – răspunsuri ce sunt adesea mai simple, în sensul lui Occam, adică *mai eficiente* decât cele mai bune dintre tentativele noastre anterioare: „Oare toate acestea ar putea fi cel puțin o parte din soluția la enigmele profunde și persistente referitoare la originile ideilor noi și la rezolvarea unor probleme în mod creativ?”[37].

Conștiința joacă în acest proces autogenerat același rol ca în percepție și în învățarea din experiență. Toate procesele cognitive conștiente au în comun faptul că presupun travaliul psihic necesar al reconsolidării – întoarcerea predicțiilor consolidate la stările de incertitudine. Iată de ce visele (care sunt o formă de rezolvare a problemelor) sunt conștiente[38].

Dar există și alte tipuri de gândire în afară de hoinăreala minții. Să analizăm un al doilea tip: reveria *intenționată*. Dacă există vreun proces cognitiv care merită rolul antagonist în opoziția tradițională dintre gândire și faptă, atunci acesta este. Vorbim aici de gândire *în loc de* acțiune, care ne inhibă impulsurile motoare în timp ce sistemul caută în imaginație soluţii la probleme în funcţie de ceea ce simte. Când abordăm acest tip de gândire, ne ajustăm preciziile astfel încât erorile de predicție motoare sunt suprimate[39]. Până la urmă, rostul gândirii deliberative este să-ți imaginezi că faci diverse lucruri ca să evaluezi *dinainte* care sunt consecințele probabile ale realizării lor în realitate. (Am folosit în capitolul 5 exemplul directorului pe care îmi venea să-l lovesc.)

Cum ne imaginăm viitorul? Cam tot așa cum ne amintim trecutul – care se dovedește a fi, mult mai frecvent decât ne place să recunoaștem, un trecut imaginat. Memoria episodică este un proces *construit* în care obiectivele și contextele curente contribuie în mare măsură la ceea ce este readus în memorie. Astfel, trecutul este retrăit în mod selectiv și tendențios în funcție de cererile curente de travaliu predictiv. Din nou, cazul domnului S. – căruia îi „lipsea un cartuș de memorie” – este un exemplu excelent, tocmai pentru că patologia lui îi exagera atât de mult mecanismele aflate la baza rememorării normale. Memoria lui episodică funcționa alarmant de mult în propriul interes.

Bineînțeles, nu există cartușe de memorie, dar sistemele neuronale implicate în călătoriile în timp la nivel mintal depind de *hipocamp*, care este crucial pentru introducerea așa-numitului „*mine-ness*”[*2] perspectival în procesele mnezice corticale în mod normal inconștiente[40]. Studiile contemporane asupra memoriei episodice relevă faptul că hipocampul este la fel de implicat în imaginarea viitorului ca şi în retrăirea trecutului[41]. Din acest motiv, David Ingvar vorbește despre „amintirea viitorului”[42], iar Daniel Schacter consideră că hipocampul – împreună cu celelalte structuri cerebrale responsabile pentru memoria episodică – susține o „simulare episodică constructivă” a viitorului, care presupune „recombinarea flexibilă a unor detalii din evenimente trecute în scenarii noi”. În opinia lui Schacter, sistemul memoriei episodice își obține valoarea adaptivă mai mult din capacitatea sa de a imagina viitorul decât din cea de rememorare a trecutului. Creierul, conchide el, este „un organ esențialmente *prospectiv*, proiectat să folosească informații din trecut și din

prezent ca să genereze predicții despre viitor”[43]. Toate acestea ar trebui să sune deja familiar.

Al treilea și ultimul tip de gândire despre care voi discuta este gândirea cu *cuvinte*. Aparent, această capacitate este caracteristica remarcabilă a cogniției umane. Oamenii descriu limbajul drept un instrument de comunicare, ceea ce și este. Dar înainte de toate e un instrument de abstractizare. Unii filosofi consideră că această „altă” funcție a limbajului ține de supracomunicare, dar eu prefer să consider că ține de precomunicare – este dificil să ne imaginăm cum ar fi putut să apară vorbirea (spre deosebire de vocalizare) în lipsa abstractizării. Limbajul nu doar exprimă gânduri pe care le avem deja, ci formează gânduri noi.

Un experiment făcut de Gary Lupyan și Emily Ward relevă această funcție a limbajului. Cei doi au folosit o tehnică numită „suprimare continuă a instantaneelor” (*continuous flash suppression*) – alt tip de percepție bistabilă, similară rivalității binoculare –, în care o imagine prezentată încontinuu unui ochi este suprimată din conștiință printr-un flux schimbător de imagini prezentate celuilalt ochi[44]. În acest experiment, subiecții vedeau cu un ochi imaginea unui obiect familiar – cum ar fi un scaun, un dovleac sau un cangur –, iar cu celălalt vedeau o serie de linii sinuoase. Liniile sinuoase au suprimat din conștiință imaginea stabilă. Totuși, chiar înainte să privească liniile sinuoase și obiectul, subiecții au auzit unul dintre următoarele trei: (1) cuvântul denotând obiectul suprimat (de exemplu, „dovleac”, când obiectul era un dovleac), (2) cuvântul denotând un obiect diferit (de exemplu, „cangur”, când obiectul era un dovleac) sau (3) doar zgomot alb. Când li s-a cerut să indice dacă au văzut ceva sau nu, posibilitatea ca

subiecții să raporteze că au văzut în mod conștient obiectul stabil atunci când cuvântul auzit se potrivea cu obiectul a fost semnificativ mai mare decât atunci când imaginea a fost însoțită de un cuvânt nepotrivit sau de zgomot alb. De fapt, auzirea cuvântului greșit a redus și mai mult șansele de a vedea obiectul suprimat. Explicația este că, „atunci când informația asociată cu etichete verbale se potrivește cu activitatea de intrare (de jos în sus), limbajul oferă un *impuls* de sus în jos percepției, propulsând în conștientizare o imagine altfel invizibilă”[45]. Cu alte cuvinte, ea crește cântărirea preciziei pentru imaginea perceptivă.

Ca să înțelegem forța acestui mecanism, să observăm că (în realitate, nu doar în acest experiment) ne furnizăm tot timpul asemenea etichetări de sus în jos, prin procesul „vorbirii interioare”. Impulsul astfel furnizat este o formă de *amorsaj* nondeclarativ. E interesant însă că amorsajul cu cuvinte are un efect substanțial mai puternic decât amorsajul cu imagini concrete. Se dovedește că o imagine valorează ceva mai puțin decât un cuvânt, ca să nu mai zic de o mie. Asta se întâmplă probabil pentru că abstracțiunile (care se află mai în profunzimea ierarhiei predictive) pot să realizeze mai mult decât imaginile când e vorba de chestiuni precum „calitatea de a fi câine”. Astfel, de exemplu, Lupyan a descoperit că auzirea cuvântului „câine” are șanse considerabil mai mari să domine suprimarea continuă a instantaneelor decât simpla auzire a unui lătrat[46]. Abstractizarea are efecte mai profunde. Și astfel, când subiecții sunt amorsați implicit să fie atenți la „vehicule” versus „oameni” în timp ce urmăresc un videoclip, amorsajul verbal modifică reglarea unor întregi populații neuronale, făcându-le să fie mai sensibile la prezența unei clase de obiecte

versus cealaltă[47]. Cuvintele au puterea de a impulsiona categorii semantice întregi. Într-adevăr, multe asemenea categorii ar fi inimaginabile – și, de aceea, imperceptibile – în lipsa etichetelor lor verbale. Tipul de concepte abstracte pe care le analizăm în cartea de față e cazul cel mai evident în care este valabilă această afirmație. A văzut vreodată cineva energia liberă? Totuși, odată ce ne gândim la „energia liberă”, îi putem vedea mecanismele peste tot.

Dacă asta pot să facă unu-două cuvinte, ce se întâmplă când începem să alăturăm sute de cuvinte? De exemplu, să ne gândim la potențialul narațiunilor personale: povestirile abstractizate pe care ni le spunem despre curgerea și sensul vieții noastre. Andy Clark scrie că „asemenea narațiuni funcționează ca elemente de nivel superior în modelele care ne structurează predicțiile despre sine și, astfel, ne influențează acțiunile și aleg erile viitoare”[48].

Bineînțeles, aceste narațiuni sunt *coconstruite* de obicei împreună cu alții, pe parcursul vieții, începând cu diada mamă – nou-născut. Aceasta introduce funcția de comunicare a limbajului. Nu e neapărat necesar ca manipularea artificială a preciziei să vină numai din modelele noastre generative – ea poate veni la fel de bine şi din modelele altora, dacă aceștia au capacități similare de abstractizare. Clark include această a doua funcție a limbajului în categoria „predicție reciprocă continuă”, iar Andreas Roepstorff și Chris Frith vorbesc de „partajarea scriptului” și de „controlul de tip *top-top* al acțiunii”[49]. Pe scurt, etichetarea abstractă ce controlează precizia unei persoane poate fi comunicată direct preciziei alteia, evitându-se învățarea laborioasă de jos în sus.

Roepstorff și Frith argumentează această idee comparând

efectele adresării de instrucțiuni verbale unor oameni care să-i ajute să îndeplinească o sarcină versus procesul anevoios al dresajului necesar pentru ca o maimuță să înregistreze suficiente informații ca să facă același lucru. În exemplul folosit de Roepstorff și Frith – un joc de sortare a unor cărți de joc –, oamenii dobândesc comportamentul vizat după câteva minute de instrucțiuni verbale, iar pentru ca maimuțele să-l deprindă, e nevoie de un an întreg de condiționare operantă. Știm din dovezile furnizate prin RMNf că oamenii și maimuțele din acest experiment au folosit regiuni cerebrale echivalente – același set cognitiv – pentru *a îndeplini* sarcina. Diferită a fost doar metoda prin care *au dobândit* comportamentul.

Limbajul înseamnă chiar mai mult. El deschide calea multor tehnici diferite de îmbunătățire a preciziei, cum sunt observațiile, teoriile și ecuațiile prezentate în această carte. Datorită cuvintelor (și altor simboluri, precum cele matematice), modelele dobândite pe parcursul unei vieți pot deveni obiecte stabile care să fie analizate și ameliorate sistematic de alții – nu numai de contemporanii noștri, ci și de generațiile viitoare. Cu ajutorul minunatelor sale grade de generalitate și de specificitate, limbajul ne permite chiar să proiectăm în conștiință ceva din structura ierarhiei predictive. Aceste ajutoare puternice ale cogniției nu sunt disponibile speciilor nonsimbolice. E foarte dificil să ne imaginăm știința, tehnologia și cultura în general în lipsa limbajului.

Eu bănuiesc că limbajul a evoluat în principal din pulsiunea JOACĂ. Am explicat în capitolul 5 cum generează JOACA formarea regulilor sociale. Regulile reglementează comportamentul de grup și, prin aceasta, ne protejează de nevoile potențial excesive ale fiecărui individ. Regula 60:40 este

o regulă socială înnăscută. Cere reciprocitate și, de aceea, facilitează dezvoltarea empatiei – capacitatea de a citi alte minți. E ușor de înțeles modul în care formarea regulilor sociale dobândite încurajează forme complexe de comunicare pentru a le exprima, precum și modul în care acest lucru, la rândul său, contribuie la apariția gândirii simbolice. Presiunea pentru a elabora reguli artificiale a crescut exponențial când oamenii au trecut de la vechile stiluri de viață de tip vânător-culegător caracteristice tuturor primatelor spre stilul de viață în așezări permanente presupunând culturi agricole și creșterea animalelor. Nu am fost pregătiți sub aspectul evoluției speciei pentru această schimbare, ce a avut loc în urmă cu doar 12.000 de ani – singura asemenea „pregătire” o constituie pulsiunea JOACĂ, esențială pentru formarea ierarhiilor sociale.

Prin urmare, este important să observăm că cortexul contribuie la JOACĂ mai mult decât la oricare dintre celelalte emoții de bază[50]. Calitatea de „ca și cum” a acestei pulsiuni este imposibil de conceput în lipsa mecanismelor corticale de tipul descris în acest capitol. JOACA ar putea fi chiar un precursor biologic al gândirii în general (adică al întregii acțiuni virtuale versus acțiunea reală) și al întregii vieți culturale.

Cititorii atenți vor observa că afirmațiile mele din această secțiune sugerează de asemenea ceva despre rolul limbajului în „terapia prin vorbire” în accepțiunea lui Freud. Ce este terapia prin vorbire dacă nu o intervenție țintită în narațiunea cuiva? În opinia mea, psihoterapia – o formă de „predicție reciprocă continuă” – este și o formă de JOACĂ. Dar acum ating teme care merită dezvoltate aparte. Voi încheia aceste reflecții despre cortex printr-o analiză a *diferenței* esențiale dintre conștiința corticală și cea a trunchiului cerebral.

Panksepp era foarte interesat de tendința noastră de a asocia anumite culori cu anumite senzații[51]. De pildă, în mod obişnuit, descriem culorile din spectrul roșu drept calde, iar pe cele din spectrul albastru drept reci. Vorbim de ceva arbitrar? Oare culorile pot fi inversate? Căldura și răceala exteroceptive nu sunt, desigur, calități vizuale, cum sunt roșul și albastrul, ci atribute ale senzației somatice, fiind însă și foarte apropiate de ceea ce Panksepp numea „afecte senzoriale”, precum dezgustul, durerea și uimirea[52]. Din mai multe puncte de vedere, aspectele hedonice ale căldurii și răcelii includ o valoare biologică – apropierea fizică de alții este caldă, sexul este cald, focul este cald etc. E posibil ca spectrele de culori asociate cu focul și gheața să-și fi căpătat semnificațiile afective în acest fel și tot așa, probabil, culorile fructelor coapte versus ale fructelor necoapte. Panksepp a speculat că aceste asocieri ar putea fi un vestigiu al originilor evolutive ale qualia perceptive, amintirea vremurilor când singurele noastre percepții conștiente erau afectele senzoriale.

Dar există o mare diferență între simțirea afectelor și atribuirea unor valențe afective obiectelor percepute la un nivel exterior, cum ar fi fructele coapte sau focurile fierbinți. Diferența este că afectele sunt înnăscute, dar atribuirea de valență obiectelor percepute externe este *dobândită*. Calitățile vizuale și alte calități perceptive ale obiectelor percepute devin „asociate” cu afectele în sensul empirist[53]. Noi atribuim valori lumii. Deși *unele* asemenea asocieri urmează regularități ce există în natură – între roșu, foc și calitatea de a fi copt, pe de o parte, și anumite senzații plăcute, pe de altă parte –, chiar și aceste asocieri sunt dobândite și, prin urmare, sunt supuse capriciilor experienței individuale.

Pe scurt, asemenea asocieri sunt *contextuale*. De exemplu, focul nu este *întotdeauna* un lucru bun, motiv pentru care, probabil, culorile din spectrul cald precum roșu uneori denotă lucruri bune, ca sexul, și alteori lucruri rele, ca pericolul. Acesta este şi motivul pentru care vorbim nu numai de sex „fierbinte”, ci și de conflict „fierbinte”. Acest principiu se aplică într-o măsură din ce în ce mai mare în cazul asocierilor tot mai individualizate care constituie iconografia subiectivă a fiecărei persoane. El funcționează în logica Lisei Feldman Barrett.

Panksepp considera că valorile afective asociate cu unele calități perceptive sunt un vestigiu al originilor lor în afectele senzoriale. Dar nu trebuie să pierdem din vedere faptul că conexiunea înnăscută se rupe când facem legături între calități perceptive precum roșeața sau căldura și trăirile erotice ori culoarea albastră sau răceala și sentimentul de tristeţe. Iată motivul pentru care este atât de ușor să condiționăm asocierile dintre culori și afecte.

Regularitățile la nivel de populație ale acestor asocieri, care într-adevăr corespund naturii până la un anumit punct, trebuie să fie baza unor forme de artă precum pictura, muzica și dansul. Le apreciem calitățile vizuale, auditive și somatosenzoriale – cel puțin parțial – prin intermediul afectelor senzoriale și al conotațiilor pe care ni le evocă. Pentru că am definit conștiința drept incertitudine resimțită, e interesant de observat și rolul surprinderii în experiența estetică. Asta se aplică și în cazul umorului: o operă de artă banală e ca un banc a cărui poantă o intuiești dinainte. Dar aceste teme sunt prea vaste ca să le tratăm corespunzător aici[54].

Putem concluziona că aceste qualia perceptive sunt diferite de cele afective în sensul că ele *nu au inerent* valențe hedonice,

ci își dobândesc valența când sunt în relație cu afectul. Lumina și undele sonore ce bombardează mereu cortexul nu sunt resimțite întotdeauna conștient[55], dar, atunci când le conștientizăm, le simțim drept *context* (vezi pp. 151-152).

Până la urmă, percepția constă din calcularea statisticii inferențiale cu privire la distribuțiile de probabilitate ale așa-numitelor *spike trains*, precum și din compararea unor asemenea probabilități, toate într-o ierarhie în serie a homeostatelor ce optimizează precizia[56]. Ierarhia generează în ochii minții noastre o reprezentare grafică a „răspunsurilor evocate de echipament” – *fenomene* în accepțiunea lui John Wheeler. Aceste fenomene constituie realitatea virtuală afișată înaintea fiecăruia (și de fiecare) dintre noi. Stările de repaus ale miliardelor de mici homeostate – toate încorporate unele în altele, de la suprafață până la niveluri profunde – reprezintă *încrederea* noastră într-un context preconizat. Contextul este ceea ce ne așteptăm să se întâmple dincolo de zona Markov, când încercăm să rezolvăm o nevoie prioritizată. Prioritizarea unei nevoi declanșează travaliu cognitiv înrădăcinat într-o credință aferentă: efectul preconizat. Ceea ce percepi conștient, prin mecanismul cu care suntem familiari de-acum, nu e acest context preconizat, ci afișarea în serie – într-o ierarhie profundă – a *devierilor* prioritizate de la așteptările tale.

La periferia ierarhiei, asemenea fluctuații survin tot timpul; e aproape imposibil să prezici prezentul în fiecare detaliu. Încrederea noastră fluctuantă în aspectele predominante ale așteptărilor noastre senzorimotoare este resimțită sub forma culorilor, tonurilor și altora asemenea – adică în aranjamente ale variabilelor categoriale exteroceptive calculate de specia noastră. Acestea includ nu numai perceperi ale lumii

exterioare, ci și devieri fluctuante de la stările proprioceptive preconizate ale avatarului ce îmi reprezintă corpul, întrucât felul cum îmi simt corpul nu e mai puțin virtual decât felul cum simt lumea exterioară[57].

Erorile de predicție a preciziei în percepție și în propriocepție (când sunt predominante) se consemnează drept qualia exteroceptive. În schimb, încrederea schimbătoare în credința fenotipică ce a generat de la început procesul senzorimotor este resimțită sub formă de afect. Indiferent ce ne propunem să realizăm, simţim în fiecare moment cât de bine ne descurcăm. Cu siguranță că şi tu simți asta acum, nu-i așa? O fluctuaţie în interiorul tău pe măsură ce citești aceste cuvinte, un flux de incertitudine oscilantă, care te-ar putea opri dacă devine prea puternică. Calitatea sa fundamentală, afectul, este resimțită pe tot parcursul vieții tale și ea reglează până la urmă tot ceea ce faci.

Aș zice că experiența noastră de zi cu zi nu constă până la urmă decât în asta.

*1 Sigmund Freud, „Dincolo de principiul plăcerii”, în *Opere esențiale*, vol. III: *Psihologia inconștientului*, traducere de G. Purdea și Vasile Dem. Zamfirescu. București: Trei, e-book (n.tr.).

*2 Faptul de a fi al meu (n.tr.).

11

Problema dificilă

Fizicianul Paul Davies scrie:

> Dintre numeroasele proprietăți complicate ale vieții, fenomenul conștiinței iese în evidență în mod special. Originile sale sunt fără îndoială cea mai dificilă problemă cu care se confruntă știința astăzi și singura ce rămâne aproape de nepătruns chiar și după două milenii și jumătate de deliberări [...]. Conștiința este problema numărul unu a științei, a existenței chiar[1].

Aș putea cita multe afirmații similare făcute de alți specialiști. „Problema dificilă” (cum e numită pe scurt și respectuos) întreabă de ce și cum anume a fost posibil ca tu – „bucuriile și tristețile tale, amintirile și ambițiile tale, sentimentul identității tale personale și al liberului-arbitru”[2], pe scurt, experiența existenței tale – să izvorăști din procesele fiziologice care au loc în celulele cerebrale. Aceste celule nu sunt fundamental diferite de cele din care sunt alcătuite alte organe. Deci cum fac ele ca „tu” să existe?

Întrebarea nu e deloc nouă, ci e probabil cel mai vechi

dintre toate misterele care îi preocupă pe oameni. Pe vremuri avea forma „Cum a ajuns sufletul meu să sălăşluiască în trupul meu?”, dar în forma sa curentă a fost exprimată în 1995 de filosoful David Chalmers. Iată celebra lui formulare:

> E incontestabil că unele organisme sunt subiecte ale experienței. Dar întrebarea cum se face că aceste sisteme sunt subiecte ale experienței ne pune în încurcătură. De ce, când sistemele noastre cognitive prelucrează informații vizuale și auditive, avem experiențe vizuale sau auditive: calitatea albastrului-închis, senzația notei do din octava a patra? Cum putem să explicăm de ce există *ceva cum e* să ai o imagine mentală sau să simţi o emoție? În general, se consideră că experiența apare de la o bază fizică, dar nu avem nicio explicație satisfăcătoare a motivelor pentru care şi a felului cum apare astfel. De ce ar trebui ca procesarea fizică să dea naștere unei vieți interioare bogate? Dintr-o perspectivă obiectivă, pare irațional să existe această necesitate și totuși ea există[3].

Formularea lui Chalmers se inspiră în mare parte din ideile unui articol al filosofului Thomas Nagel, „What is it like to be a bat?” („Cum e să fii liliac?”, 1974). Nagel a subliniat acel *ceva-cum-e* al experienței subiective. A arătat că „un organism are stări mintale conștiente dacă și numai dacă există ceva cum e să *fii* acel organism – ceva cum e *pentru* acel organism”, adăugând: „Dacă admitem că o teorie fizică a minții trebuie să explice caracterul subiectiv al experienței, atunci trebuie să acceptăm că nicio concepţie disponibilă în prezent nu ne oferă vreun indiciu despre cum s-ar putea face acest lucru”. Nagel a

concluzionat: „Pare improbabil să putem lua în calcul vreo teorie fizică a minții până nu analizăm mai mult problema generală a ce înseamnă subiectiv și obiectiv”.

Două decenii mai târziu, Chalmers a reafirmat „problema generală” în termenii respectivi pentru că specialiștii în științele creierului tocmai începuseră să analizeze conștiința la un nivel experimental. Datorită progreselor tehnologice pe care le-am descris în capitolul 1, ei credeau că acum se poate răspunde la întrebări precum „În ce mod transformă celulele cerebrale în experiențe procesele fiziologice?”. Unul dintre primii cercetători care au abordat problema la nivel experimental a fost sir Francis Crick, specialist în biologie moleculară și membru al echipei ce a descoperit structura ADN-ului. El a vorbit despre asta în volumul *The Astonishing Hypothesis: The Scientific Search for the Soul*, pe care l-a publicat cu doar un an înainte ca Chalmers să-și declare dubiile. Iată ce a scris Crick:

> Ipoteza uimitoare este că tu, bucuriile și tristețile tale, amintirile și ambițiile tale, sentimentul identității tale personale și al liberului-arbitru, nu ești *de fapt* decât comportamentul unui mare ansamblu de celule nervoase, cu moleculele lor aferente[4].

Chalmers nu a contestat categoric ipoteza că conștiința are o bază fizică. După cum tocmai am spus, el însuși a afirmat că, „în general, se consideră că experiența apare de la o bază fizică”. A spus doar că „nu avem *nicio explicație satisfăcătoare* a motivelor pentru care şi a felului cum apare astfel”. Prin aceasta a reluat aserțiunea lui Nigel că „nicio concepţie

disponibilă în prezent nu ne oferă *vreun indiciu* despre cum s-ar putea face acest lucru". Crick afirmase că se poate găsi o explicație satisfăcătoare: că tehnologia disponibilă în prezent ar putea să identifice cu ușurință ceea ce numea „corelativele neuronale ale conștiinței". Credea că, dacă izolăm părțile anatomice ale creierului ce sunt necesare pentru conștiință, precum și funcțiile lor fiziologice specifice, am putea rezolva în mod științific problema minte-corp. Crick ne-a recomandat să începem căutarea concentrându-ne numai asupra unuia dintre corelativele neuronale ale conștiinței: procesele cerebrale care disting *vederea* conștientă de cea inconștientă. Probabil că de aici am putea extrapola la restul conștiinței. Părea destul de logic. Sigur există un corelativ neuronal al experienței vizuale, nu-i așa?

După cum am văzut, deși numai vederea *corticală* este conștientă, cortexul poate să proceseze stimuli vizuali și în mod inconștient, la fel și coliculii superiori. Prin urmare, conform lui Crick, problema conștiinței vizuale este o întrebare simplă referitoare la ce se petrece în cortexul vizual când prelucrează informații *conștient*, dar nu se petrece când o face *inconștient*, ceva ce nu se întâmplă nici în coliculii superiori. Din motivele prezentate mai sus, eu cred că a pornit cu stângul, ca să vorbesc dintr-o perspectivă anatomică; ar fi trebuit să se axeze pe trunchiul cerebral, nu pe cortex, pe afect, nu pe vedere. Rezervele lui Chalmers au fost însă mai importante.

Perspectiva lui Crick – care a devenit abordarea principală în neuroștiințele cognitive – eludează ceea ce Chalmers numește partea „dificilă" a problemei minte-corp. Izolarea corelativelor neuronale ale conștiinței este partea „ușoară". Nu face decât să identifice procesele cerebrale specifice care *se*

corelează cu experiența și nu explică cum o *cauzează*. Asta este partea dificilă a problemei: cum și de ce *produc* activitățile neurofiziologice experiența conștiinței?[5] Cu alte cuvinte, cum anume *devine* materia minte? Conform lui Chalmers, deși noi, specialiștii în neuroștiințe, putem să explicăm cum se prelucrează în creier informațiile neuronale în timp ce avem experiențe vizuale, asta nu explică modul în care procesele cerebrale respective *se transformă în* experiență. După cum s-a exprimat John Searle, alt mare specialist în filosofia minții: „Cum anume depășește creierul obstacolul dintre electrochimie și sentimente?”[6]. Această întrebare ne derutează la fel de mult şi când este formulată invers: cum se transformă lucrurile imateriale precum gândurile și sentimentele (de exemplu, te hotărăști să-ți faci un ceai) în acțiuni fizice ca prepararea unui ceai?[7]

Amploarea acestei lacune explicative[8], cum o numesc filosofii, este bine ilustrată de „argumentul cunoașterii”, care sună cam așa[9]. Imaginează-ți o specialistă în neuroștiințe oarbă congenital pe nume Mary, care știe tot ce se poate ști despre corelativele neuronale ale vederii. Deși ea poate să explice toate realitățile fizice ale prelucrării informațiilor vizuale – până la nivelul celular, inclusiv impactul undelor de lumină asupra conurilor și bastonașelor fotosensibile, precum și modul în care aceste unde sunt transformate în impulsuri nervoase, apoi aceste impulsuri sunt propagate prin corpul geniculat lateral spre cortex, apoi acestea sunt prelucrate acolo mai departe de coloane de neuroni bine ordonați, organizați în număr mare într-o diversitate de module de prelucrare a informațiilor răspândite peste tot în scoarța cerebrală, ale cărei multiple fluxuri de procesare vizuală specializată sunt bine înțelese –, ea

tot nu ar ști *cum e* să ai experienţa vederii. Fiind oarbă din naștere, ea nu ar ști nimic despre calitățile resimțite a ceea ce înseamnă roșu și ce înseamnă albastru – de exemplu –, care sunt, până la urmă, materia propriu-zisă a vederii conștiente. Asta nu numai din cauză că ea nu a avut niciodată experienţa unor asemenea calități, ci și din cauză că nimic din *cunoștințele* ei de anatomie și de fiziologie despre corelativele neuronale ale vederii nu explică cum e să vezi. Dacă ar căpăta brusc capacitatea de a vedea, Mary ar învăța ceva cu totul nou despre vedere – ceva pentru care nu a pregătit-o nimic din înțelegerea ei mecanicistă. *Prin urmare, realitățile fizice nu explică de ce și cum există ceva cum e să vezi.* Ele explică numai de ce și cum decodifică creierul informații vizuale: cum vede *el*, și nu cum vezi *tu*. Această așa-zisă ireductibilitate la baza fizică a ceea ce filosofii numesc „qualia” – acel ceva-cum-e al experienței subiective – este problema dificilă. Conform lui Chalmers, acesta este „misterul central al conștiinței”[10].

De-a lungul timpului, această ireductibilitate percepută i-a făcut pe marii gânditori, nu în ultimul rând pe fizicianul-filosof John Locke, să tragă concluzia că experiențele conștiente *nu fac parte din universul fizic*[11]. Întrucât e clar însă că qualia experienței există, acești gânditori le-au relegat într-o dimensiune *nonfizică* a realității, pe care (mulți dintre ei) au descris-o drept „epifenomenalistă”[12]. Locke a fost cel care a subliniat în „argumentaţia spectrelor inversate” că este posibil din punct de vedere logic ca o persoană să experimenteze calitatea albăstrimii la fel cum eu experimentez calitatea roșeții și să o numească tot „roșu”, ca mine, deşi ceea ce vede în mod conștient e ceea ce eu aș numi „albastru”. Consecința acestei afirmații e că așa ceva *nu schimbă lucrurile* – aceste două

extreme ale spectrului de culori ar putea fi inversate imediat în relație cu qualia lor resimțite. Așadar, conform lui Locke, conștiința nu este explicată de corelativele ei fizice – relația dintre mecanismul fizic (în acest caz, lungimile de undă relative ale luminii) și qualia psihologice nu este una cauzală.

Astfel, conștiința devine ceva ce doar rezultă pe lângă lanțul de evenimente fizice din creier – un fel de produs secundar –, fără impact asupra structurii cauzale. Locke scria în secolul al XVII-lea, dar această perspectivă nu este arhaică, ci e larg răspândită, nu numai în rândul filosofilor. După cum au afirmat recent doi specialiști reputați în științe cognitive: „Conștientizarea personală este asemenea curcubeului ce însoțește procese fizice din atmosferă, dar nu exercită nicio influență asupra lor”[13].

Dacă experiențele conștiente nu joacă niciun rol în funcționarea lumii fizice, de ce mai există? Ce anume adaugă conștientizarea la prelucrarea informațiilor ce are loc oricum, în mod inconștient? Ce rost are să devii conștient de procesele cerebrale, dacă conștientizarea nu are nicio influență asupra lor?[14]

Această afirmație nu e valabilă numai în cazul vederii. Una dintre dovezile experimentale cele mai convingătoare ale opiniei conform căreia conștiința e un simplu epifenomen al proceselor cerebrale este o observație a lui Benjamin Libet, care spune că decizia ta subiectivă de a iniția o mișcare este precedată (cu aproximativ 300 de milisecunde) de unde cerebrale măsurabile ce anunță declanșarea mișcării pe care crezi că ai inițiat-o tu[15]. Cu alte cuvinte, inițierea fizică a mișcării (în creier) începe înainte să te hotărăști în mod conștient să te miști. De fapt, nu „tu” o inițiezi. În general, se

consideră că această descoperire demonstrează că alegerea conștientă – „liberul-arbitru” – este o iluzie. Dacă nu există liber-arbitru, atunci ce mai rămâne din conștiință? După cum am văzut, răspunsul lui Crick este că tu, bucuriile și tristețile tale etc., nu exiști „de fapt”. Dacă am avea o explicație fizică adecvată pentru asemenea qualia, spune Crick, le-am putea *suprima*.

Mi s-a părut întotdeauna imposibil să accept atât argumentul conform căruia qualia conștiente *există într-un univers paralel*, cât şi argumentul că ele *nu există deloc*. Și așa ar trebui să faci și tu, pentru că tu ești conștiința ta. Este, evident, absurd să pretinzi că ești ca un curcubeu care nu exercită nicio influență asupra corpului tău fizic. La fel de absurd e să pretinzi că „tu” nu ești, de fapt, decât comportamentul unor neuroni, deci nu exiști cu adevărat. Aceste afirmații sunt contrazise de fiecare moment al experienței tale.

Afirmația că „tu” nu exiști contrazice clar și una dintre cele mai celebre concluzii din întreaga gândire filosofică. După o viață întreagă în care a meditat la *ce anume înseamnă certitudinea absolută*, René Descartes (în „filosofia îndoielii”) a ajuns la celebra concluzie că *singurul* lucru în legătură cu care nu trebuie să avem nicio îndoială este faptul că existăm: „Cuget, deci exist”. Cu alte cuvinte: ai experienţe, deci exiști.

Aceasta este problema dificilă a lui Chalmers: întrucât există în mod clar un sine ce are experiențe, cum îl integrăm în concepția noastră fizică despre univers?

Eu nu sunt filosof. După cum ai văzut, interesul meu pentru neurologia conștiinței s-a născut și s-a dezvoltat independent de literatura filosofică și mărturisesc că o bună parte din câtă

filosofie am citit mai degrabă mă derutează. Cu toate acestea, în cartea de față încerc să ofer un răspuns științific firesc la „problema dificilă”. Nu am pretenția că am risipit orice mister din metafizica conștiinței. După cum vom vedea, rămân încă întrebări profunde. Dar cred că am demonstrat că procesele naturale, ce se desfășoară în funcție de diferitele lor necesități, pot (pe parcursul evoluţiei) să genereze ceva foarte asemănător cu lumile particulare ale experienței pe care le avem.

Ce relevanță are asta pentru problemă, așa cum o formulează Chalmers? Cum rezolvă explicația pe care o dau lacuna explicativă dintre lumile interioare și cele exterioare?

Să pornim de la o observație simplă: legile ce guvernează funcțiile psihice precum percepția, memoria și limbajul sunt *abstrase* atât din date subiective, cât și din date obiective. Ca să dau un exemplu, legea lui Ribot, o abstractizare științifică despre memoria pe termen lung, explică faptul observabil că memoria resimțită *intern* a ceea ce s-a întâmplat acum zece ani este consolidată mai sigur decât ce s-a întâmplat acum zece minute. Acesta este motivul pentru care persoanele în vârstă tind să uite mai degrabă evenimente recente decât evenimente petrecute cu mult timp în urmă. Afirmația e valabilă și în cazul urmelor mnezice fiziologice, observate *extern*: cele vechi de zece ani sunt mai bine consolidate decât cele vechi de zece minute (în exact aceeași măsură)[16]. Așadar, gradientul temporal din legea lui Ribot nu este psihologic sau fiziologic, ci de ambele feluri. De asemenea, să ne gândim la legea lui Miller, o abstractizare științifică despre memoria pe termen scurt, ce explică limitarea resimțită a capacității de a ține minte: în orice moment, nu poți reține decât șapte unități de informație (plus/minus două). Şi durata amintirilor pe termen scurt poate

fi măsurată: de obicei, ele durează 15-30 de secunde[17]. Exact aceeași limitare a capacității poate fi observată la nivel fiziologic: presupune depleționea neurotransmițătorilor[18]. Prin urmare, legea lui Miller, la fel ca şi legea lui Ribot, este atât psihologică, cât și fiziologică.

Legi ca acestea, care, în principiu, pot fi cuantificate și deci exprimate matematic, explică manifestările duale ale unei funcții abstrase numite „memorie”. Cu siguranță același lucru e valabil și în cazul „conștiinței”. Dacă fenomenele conștiinței sunt lucruri naturale (ce altceva ar putea fi?), atunci și ele trebuie să fie reductibile la niște legi.

Această concluzie nu este radicală sau idiosincrazică. Întreaga știință cognitivă, care domină de la sfârșitul secolului XX domeniul în care m-am specializat, se întemeiază pe asta. Să ne amintim că în capitolul 1 am scris:

> Cel mai interesant [la „prelucrarea cognitivă a informațiilor”] e că poate fi implementată cu tipuri foarte diferite de echipamente fizice. Astfel, reușim să vedem natura fizică a minții într-o lumină nouă. Ni se sugerează că mintea (concepută ca prelucrare a informațiilor) este o *funcție*, și nu o structură. Din această perspectivă, funcțiile de „software” ale minții sunt implementate de structurile „hardware” ale creierului, dar aceleași funcții pot fi implementate la fel de bine şi de alte substraturi, precum calculatoarele. Astfel, atât creierii, cât și calculatoarele îndeplinesc funcții de *memorie* (codează și stochează informații) și de *percepție* (clasifică tipare ale informațiilor de input prin compararea lor cu informațiile stocate), precum și funcții de *execuție* (execută decizii referitoare la

ce este de făcut ca reacție la asemenea informații).

Totuși, am continuat imediat:

> Aceasta este puterea a ceea ce a ajuns să se numească abordarea „funcționalistă”, dar este și punctul ei slab. Dacă aceleași funcții pot fi îndeplinite şi de calculatoare, despre care presupunem că nu sunt ființe conștiente, atunci se justifică să reducem mintea la simpla prelucrare a unor informații?

Și așa ajungem la esența problemei dificile. Conform lui Chalmers, *problema dificilă a conștiinței nu poate fi rezolvată prin explicații „funcționaliste”*:

> Problemele simple sunt simple tocmai pentru că se referă la explicarea abilităților și funcțiilor cognitive. Ca să explicăm o funcție cognitivă, nu trebuie decât să specificăm un mecanism care o poate îndeplini. Metodele științelor cognitive se potrivesc pentru acest tip de explicație, deci sunt adecvate pentru problemele simple ale conștiinței. În schimb, problema dificilă este dificilă tocmai pentru că nu se referă la îndeplinirea unor funcții. Problema persistă chiar și atunci când este explicată îndeplinirea tuturor funcțiilor relevante [...]. Problema dificilă este dificilă și aproape unică pentru că trece *dincolo de* problemele legate de îndeplinirea unor funcții. Ca să înțelegem acest lucru, să observăm că, și după ce am explicat îndeplinirea tuturor funcțiilor cognitive și comportamentale în vecinătatea experienței [...], tot mai poate să rămână o întrebare fără răspuns: *de ce îndeplinirea*

> *acestor funcții este însoțită de experiență?* O explicație simplă a funcțiilor lasă deschisă această întrebare [...]. De ce toată această prelucrare a informațiilor nu continuă „pe întuneric", fără să simţim nimic în interior?[19]

În capitolul 4, am rezumat tentativele eșuate de specificare a unui mecanism care să poată îndeplini funcția conștiinței. Răspunsul lui Chalmers la prima dintre ele ne lămurește cu privire la atitudinea lui față de toate:

> Crick și Koch sugerează că oscilațiile [gama sincronizate] sunt *corelativele* neuronale ale experienței. Această afirmație e discutabilă [...], însă, chiar dacă este acceptată, rămâne întrebarea legată de *explicație*: de ce oscilațiile creează experiență? Singura bază pentru o conexiune explicativă este rolul lor în fixare și stocare, dar întrebarea de ce fixarea și stocarea ar trebui să fie însoțite de experiență nu e discutată niciodată. Dacă nu știm de ce fixarea și stocarea ar trebui să creeze experiență, nu ne ajută să spunem o poveste despre oscilații. Și invers, dacă *am ști* de ce fixarea și stocarea creează experiență, detaliile neurofiziologice nu ar fi decât cireașa de pe tort. Teoria lui Crick și Koch prinde pentru că *presupune* o conexiune între fixare și experiență, iar astfel nu poate face nimic ca să explice legătura[20].

Iată critica lui Chalmers la a doua descriere funcționalistă importantă a conștiinței, teoria „spațiului de lucru global" a lui Newman și Baars (1993):

Am putea presupune că, potrivit lui Baars, conținutul experienței este exact conținutul spațiului de lucru. Însă, chiar dacă este așa, nimic din această teorie nu *explică* de ce sunt experiate informațiile din spațiul de lucru global. În cel mai bun caz, teoria poate să spună că informațiile sunt experiate deoarece sunt *accesibile global*. Dar acum întrebarea apare într-o formă diferită: de ce accesibilitatea globală trebuie să ducă la experiență conștientă? Ca de fiecare dată, această întrebare de legătură rămâne fără răspuns[21].

Chalmers concluzionează:

Metodele explicative obișnuite ale științelor cognitive și neuroștiințelor nu sunt suficiente. Ele au fost elaborate tocmai ca să explice îndeplinirea funcțiilor cognitive și își fac bine treaba. Dar, aşa cum se prezintă, ele sunt capabile *numai* să explice îndeplinirea funcțiilor. Când e vorba de problema dificilă, abordarea standard nu are nimic de spus [...]. Ca să explicăm experiența, avem nevoie de o abordare nouă[22].

Trebuie să clarific ce spune Chalmers aici. El afirmă că metodele științelor cognitive pot să explice numai îndeplinirea funcțiilor cognitive și că își fac bine treaba. Prin „funcții cognitive” se referă la lucruri precum percepția, memoria și limbajul. Așadar, el spune că *conștiința nu este o funcție cognitivă*. De ce spune asta? Răspunsul este: pentru că nu vorbește despre o *funcție* abstrasă numită „conștiință”, ci despre *experiența* conștiinței – despre *cum e* să percepi sau să-ți

amintești.

Problema dificilă ar fi banală dacă s-ar reduce la faptul că experiența ta individuală nu este identică cu experiența umană în general[23]. Dacă asta ar fi problema dificilă, atunci, ca să o rezolvăm, nu ar trebui decât să facem o medie a experiențelor cu martor unic ale mai multor indivizi, să le găsim numitorul comun și să explicăm *ce obținem* în termeni funcționali. Psihologii fac tot tipul asemenea lucruri. Asta am făcut și eu când am studiat confabulația: am început cu experiențele subiective ale unui individ, domnul S., le-am generalizat prin studierea experiențelor echivalente ale altor pacienți ca el, apoi am abstras numitorul comun. Această abordare a dezvăluit un principiu funcțional cu privire la confabulație, şi anume că îi face pe pacienți să se simtă mai bine; confabulația are o funcție „aspirațională”.

Dar Chalmers nu spune doar că fenomenele subiective sunt conectate cu un singur punct de vedere. El întreabă: de ce ar trebui ca o funcție, indiferent care, să fie însoțită de experiență? Și scrie:

> Cum se face că, atunci când undele electromagnetice ajung la retină și sunt discriminate și clasificate de un sistem vizual, această [funcție de] discriminare și de clasificare este resimțită ca o senzație de roșu-aprins? Știm că experiența conștientă *chiar* apare când aceste funcții sunt îndeplinite, dar însuși faptul că apare este marele mister. Există o *lacună explicativă* (sintagmă pe care i-o datorăm lui Levine, 1983) între funcții și experiență și avem nevoie de o punte explicativă ca s-o depășim. O descriere simplă a funcțiilor rămâne de o parte a lacunei, așa că materialele pentru punte

trebuie căutate în altă parte[24].

Merită să clarificăm ce trebuie să realizeze această „punte explicativă”, ca să nu pierdem din vedere ce ne interesează de fapt. Într-un articol publicat la scurt timp după ce eu și Friston l-am predat pe al nostru (2018), trei colegi cehi au anunțat că e iminentă o soluție la „problema dificilă”, prezicând că va veni de la principiul energiei libere:

> Nu există o perspectivă clară sau un consens general între filosofi și specialiștii în neuroștiințe asupra funcției conștiinței. Acesta este unul dintre motivele principale pentru care conștiința încă e o problemă atât de greu de deslușit, ce își are rădăcinile în faptul că nu a fost formulată clar niciodată o funcție a conștiinței în baza unei teorii unitare a creierului, care să o și susțină. O asemenea teorie unitară apare sub forma cadrului de codificare predictivă [...] bazat pe ideile lui Hermann von Helmholtz, conform cărora creierul este în principal un aparat pentru inferențe predictive[25].

Totuși, acești autori au adăugat un avertisment: „Din păcate, problema dificilă a conștiinței probabil nu va dispărea niciodată cu totul, din cauză că va avea mereu susținători devotați”. Și au adăugat:

> Considerăm că sunt slabe șanse ca susținătorii cei mai devotați ai problemei dificile să fie cu totul mulțumiți de [orice] concluzii ale științei empirice, din moment ce, de la bun început, esența disputei legate de problema dificilă

vizează încercările științei empirice.

Bineînțeles, când eu și Friston am primit primele evaluări din partea *Journal of Consciousness Studies*, exact asta ni s-a transmis despre articolul nostru: „Problema dificilă (conform lui Chalmers) este o problemă metafizică și, ca atare, nu se pretează la a fi «rezolvată»”. (Friston mi-a scris ulterior: „Am impresia că problema dificilă nu există ca să fie rezolvată, ci ca să fie venerată”[26].)

Ca să fiu corect faţă de Chalmers, el nu este responsabil pentru prejudecățile filosofice care îi preocupau pe colegii noștri cehi. Să ne gândim la ce spune el în fraza cu care își încheie articolul din 1995: „Problema dificilă este o problemă dificilă, dar nu există motive să credem că va rămâne veșnic nerezolvată”.

Așadar, să spunem clar ce consideră el că trebuie să realizeze puntea explicativă. Voi cita chiar criteriile lui Chalmers. Când a demontat prima teorie funcționalistă, a scris: „Nu e abordată niciodată întrebarea de ce fixarea și stocarea [sincronizate în frecvență gama] ar trebui să fie însoțite de experiență”. Așadar, când analizezi dacă lacuna explicativă este eliminată prin vreo teorie funcționalistă a conștiinței, trebuie să te întrebi: *este discutată întrebarea de ce funcția XYZ ar trebui să fie însoțită de experiență?* În ce privește a doua teorie pe care a respins-o, Chalmers a adăugat: „De ce accesibilitatea globală ar trebui să dea naștere experienței conștiente? Ca întotdeauna, această întrebare de legătură nu primește răspuns”. Așadar, când evaluezi orice altă teorie, trebuie să te întrebi: *se răspunde la întrebarea de ce funcția XYZ ar trebui să genereze experiență conștientă?* Și se pare că asta-i tot.

Dar, dacă asta-i tot, atunci de ce adaugă Chalmers că „o simplă descriere a funcțiilor [experienței] rămâne de o parte a lacunei, așa că materialele pentru punte trebuie căutate în altă parte”? De ce trebuie ca o descriere a funcțiilor să nu reuşească să elimine lacuna? Dacă recitim ce spune el în acest sens, după cum l-am citat mai sus, devine clar că el combină *două tipuri diferite de lacună explicativă*. Așadar, trebuie să spunem clar despre ce lacună vorbim, ca s-o putem elimina.

Prima este o lacună explicativă între semnalele propagate de la retină și senzațiile de roșu-aprins – adică între evenimentele fiziologice și cele psihologice. Trebuie să evidențiez că *ambele tipuri de evenimente sunt, de fapt, experimentabile*. Poți trăi evenimente fiziologice precum observarea semnalelor propagate de la retină la fel de simplu cum poți trăi evenimente psihologice precum perceperea culorii roșu-aprins. Niciunul dintre aceste evenimente nu poate fi explicat – și cu atât mai puțin suprimat – de celălalt. *Același lucru poate fi observat în două moduri*. Atunci când am experienţa subiectivă că exist, este cevaul psihic care există diferit de Mark Solms corporal pe care îl văd în oglindă? Și este Mark Solms din oglindă ceva diferit de acel Mark Solms anatomic văzut printr-un aparat de RMN? Nu înțeleg de ce atât de mulți oameni (chiar filosofi) vorbesc despre creier ca și cum ar fi cumva exceptat de la realitate *așa cum o experimentăm*[27]. Asta nu poate fi decât pentru că, atunci când spun „creierul”, ei se referă la altceva. Nu se referă la creierul pe care îl vezi și îl atingi, ci la ceva *abstras* din experienţa pe care o avem despre el – ceva de natura unui sistem *funcțional*.

Așadar, prima lacună explicativă este între *două tipuri*

diferite de experiență asociate cu două perspective observaționale diferite[28]. Acest lucru e comparabil cu *a auzi* tunete cu urechile și *a vedea* fulgere cu ochii. Lumea nu spune: „Se consideră în general că tunetul vine din fulger, dar nu avem o explicație satisfăcătoare a motivelor pentru care şi a felului cum se întâmplă așa”. Asta e pentru că lumea nu crede că fulgerul *produce* tunetul așa cum ficatul produce bilă. Se acceptă că ele sunt două manifestări ale aceluiași lucru[29]. Asta se aplică şi modurilor diferite de experiere a prelucrării informațiilor vizuale: din exterior sau din interior. Din exterior (dacă ești un cercetător dotat cu echipamentul adecvat) vezi semnale propagate de la retină[30]; din interior vezi roșu-aprins.

A doua lacună explicativă este între experiențe (de ambele tipuri) și cauzele lor fundamentale. E lacuna dintre ceea ce poți experimenta, precum nuanțele de roșu-aprins și imaginile unor neuroni activați obținute prin metodele optogeneticii, și ceea ce nu poți experimenta, precum câmpurile cuantice. Pe scurt, e vorba de o lacună între perspectiva persoanei întâi și perspectiva persoanei a treia. Să adopt o perspectivă la persoana a treia asupra propriei experiențe înseamnă să mă abstrag din experiență, deci să nu o mai am. Această perspectivă nu vizează nici creierul așa cum arată, nici mintea așa cum se simte, ci forțele care *explică* de ce și cum arată și se simte așa cum arată și se simte. Aceasta este perspectiva pe care am adoptat-o în cartea de față.

Nu sunt sigur că este evident pentru toată lumea că Chalmers, în citatul de mai sus, se referă simultan la ambele lacune. Asta poate crea confuzie. El combină lacuna dintre două tipuri de experiențe (observații extrospective versus observații introspective) cu lacuna dintre experiență în general și

mecanismele ei funcționale fundamentale. Asta complică și mai mult situația.

Mie mi se pare că majoritatea oamenilor presupun că ceea ce Chalmers numește „funcție” este sinonim cu ceea ce numește „ceea ce e fizic” și că „experiență” este, prin urmare, întotdeauna și exclusiv ceva nonfizic. Dar, dacă „ceea ce este fizic” înseamnă corpul observabil și organele sale, inclusiv creierul, atunci nu este experiat mai puțin decât ceea ce este psihologic[*1]. Mai mult, experiențele psihologice, când sunt abstrase, dezvăluie mecanismul funcțional al unor asemenea experiențe, dând naștere unor legi psihologice.

Același lucru e valabil și în cazul legilor fiziologice: și ele sunt abstrase de la experiență – de la datele fiziologice observabile. Aceste două tipuri de legi, fiind făcute din aceeași materie (explicativă) abstrasă – funcțiile –, nu sunt la fel de greu de redus una la cealaltă cum sunt cele două moduri de experienţă ce constituie categorii distincte. Astfel, legile lui Miller și Ribot sunt atât psihologice, cât și fiziologice și, din acest motiv, pot fi reduse la niște ecuații unificatoare. Faptul că nu sunt explicitate etapele intermediare dintre datele ce rezultă din experiență și mecanismele explicative complică problema dificilă și o face să pară mai dificilă decât este. Nu e nevoie să-ți imaginezi cum un tip de experiență ar putea produce alt tip de experiență: trebuie doar să găsești o teorie mecanicistă care explică ambele seturi de fenomene, indiferent de modalitatea ei perspectivală. Apoi poți să testezi predicțiile pe care le generează.

Chalmers admite că problema dificilă poate fi rezolvată. De fapt, el stabilește trei principii pe care trebuie să se bazeze

soluția. Înainte să prezint aceste principii (pe unul dintre ele trebuie să-l păstrăm pentru capitolul 12), trebuie să precizez că Chalmers consideră că nu vom putea rezolva niciodată problema dificilă dacă încercăm să o facem *reducând* „experiența” la ceea ce el numește „procese fizice”. Chalmers scrie:

> Suntem deja în poziția din care putem să înțelegem anumite realități esențiale despre relația dintre procesele fizice și experiență, precum și despre regularitățile care le conectează. *Odată ce lăsăm deoparte explicația reductivă*, putem să punem pe masă realitățile respective pentru ca ele să-și joace rolul corespunzător, de piese inițiale într-o teorie *nonreductivă* a conștiinței și de constrângeri asupra legilor de bază care constituie o teorie finală[31].

Pentru mine explicația „nonreductivă” este bună dacă înseamnă că putem renunța la sarcina imposibilă de a reduce fenomenele psihologice la unele fiziologice sau invers. Fenomenele psihologice nu pot fi reduse la unele fiziologice, la fel cum fulgerul nu poate fi redus la tunet. Fulgerul nu *cauzează* tunetul; cele două fenomene *se corelează*. Aceasta este problema simplă. Prin urmare, trebuie să reducem ambele fenomene la mecanismele lor, astfel încât să putem reduce aceste mecanisme la un numitor comun, fără să încălcăm legile fizicii. Aceasta este problema dificilă.

Totuși, Chalmers pare să se refere la ceva diferit prin termenul explicație „nonreductivă”. Pentru el, asta înseamnă că nu putem să reducem fenomenele psihologice trăite la nişte legi funcționale și atât.

Chalmers continuă:

> O teorie nonreductivă a conștiinței va consta dintr-un număr de *principii psihofizice*, care conectează proprietățile proceselor fizice la proprietățile experienței. Putem considera că aceste principii includ modul cum apare experiența din ce e fizic. În cele din urmă, aceste principii ar trebui să ne spună ce tip de sisteme fizice vor avea experiențe asociate, iar în cazul sistemelor care au, ar trebui să ne spună ce tip de proprietăți fizice sunt relevante pentru apariția experienței, precum și exact la ce tip de experiență ar trebui să ne așteptăm să genereze orice sistem fizic dat. Asta e greu de făcut.

Acesta este primul lui principiu. Îl numește „principiul coerenței structurale”. Acesta presupune coerență structurală între „proprietățile experienței” și „proprietățile proceselor fizice”. Acest lucru pare să se refere la două clase de *fenomene* care, în opinia mea – ca și a lui –, nu pot fi reduse direct una la cealaltă. Dar Chalmers vorbește și de „tip[uri] de sisteme fizice” și de „experiențele lor asociate”. Asta pare să se refere la relația dintre sistemele *funcționale* în general și experiențe în general (prin care cele din urmă „apar din” cele dintâi sau cele dintâi „le generează” pe cele din urmă). Din acest motiv, pentru Chalmers, principiul coerenței structurale exprimă modul în care „experiența apare din ce este fizic”.

Din nou, el combină fenomenele fizice și cauzele „fizice” (adică funcționale). Dacă nu deslușim etapele intermediare, vom concluziona cu siguranță că experiența *nu* apare din ce este fizic. Aceeași semnificație combinată a „ceea ce este fizic” se poate vedea şi în teoriile funcționaliste tipice pe care le-am

menționat mai sus. Ele căutau cheia experienței psihologice în mecanismul *fiziologic* X, Y sau Z: sincronismul oscilațiilor gama, fixarea și stocarea activității integrate senzoriale și frontoparietale, activarea cortexului de către nucleii intralaminari ai talamusului, „ansele de reintrare" talamocorticale etc. Folosind analogia cu tunetul și fulgerul, această combinare impune să explicăm fenomenele auditive cu ajutorul mecanismelor funcționale ale vederii. Numai în acest sens are Chalmers dreptate să spună că „o descriere simplă a funcțiilor rămâne de o parte a lacunei, așa că materialele pentru punte trebuie căutate în altă parte".

Pentru a explica psihologia în relație cu fiziologia, trebuie să ne abstragem din fenomenele observate de *ambele tipuri* (adică trebuie să inferăm mecanismele funcționale, de ambele tipuri) – apoi să ne abstragem din cele două seturi de abstracțiuni ca să vedem numitorul comun unificator. Când procedăm astfel, trebuie să ne luăm *aceeași distanță* față de ambele (adică să inferăm mecanisme suficient de profunde pentru a explica atât funcțiile psihologiei, cât și pe ale fiziologiei); numai după aceea putem să compatibilizăm fenomenele și mecanismele lor de bază.

Nu vreau să dau prea multe șabloane. Știința adevărată evoluează pe căi mai puțin ordonate. Pornim de la o idee sau o descoperire undeva, în cascada cauzalității, apoi o completăm cu ce lipsește. Dar, dacă sărim peste etapele intermediare, ajungem la cerința soluțiilor „nonreductive" la problema dificilă care sunt imposibil de obținut. Să ne amintim primele întrebări ale lui Chalmers: de ce ar trebui ca prelucrarea fizică *să dea naștere* unei vieți interioare bogate? Cum și de ce *produc* activitățile neurofiziologice experiența conștiinței? La aceste

întrebări nu se poate răspunde într-un mod nonreductiv, din cauza felului în care sunt formulate.

Prin urmare, este de înțeles concluzia lui Chalmers că o soluție reductivă, în sine, este imposibilă:

> Sugerez că o teorie a conștiinței trebuie să considere experiența fundamentală. Știm că o teorie a conștiinței necesită adăugarea a *ceva* fundamental la ontologia noastră, căci totul în teoria fizică este compatibil cu absența conștiinței [...]. O teorie nonreductivă a experienței va specifica principiile de bază care ne spun cum anume depinde experiența de caracteristicile fizice ale lumii. Aceste principii *psihofizice* nu vor interfera cu legile fizice, căci se pare că legile fizice deja formează un sistem închis. Mai degrabă, vor fi o completare la o teorie fizică. O teorie fizică dă o teorie a proceselor fizice, iar o teorie psihofizică ne spune cum anume creează procesele respective experiența. Știm că experiența depinde de procesele fizice, dar știm și că această dependență nu poate fi derivată numai din legile fizice. Noile principii de bază postulate de o teorie nonreductivă ne oferă ingredientul suplimentar de care avem nevoie ca să construim o punte explicativă [...]. Această poziție se califică drept o variantă de dualism, întrucât postulează proprietăți de bază ce sunt mai mult decât proprietățile invocate de fizică. Dar e o variantă inocentă de dualism.

Și așa, în cele din urmă, Chalmers spune ceva cam sinistru. Deoarece crede că „experiența” nu poate fi redusă la „ceea ce e fizic” (în accepțiunea lui, însemnând atât „fiziologic”, cât și

„funcțional”), este obligat să concluzioneze că *experiența nu face parte din universul fizic cunoscut*. Acesta ar părea să fie dualism pur, de tipul nu-chiar-așa-de-inocent moștenit de la Descartes și Locke[32]. Iată de ce Chalmers pretinde că conștiința necesită adăugarea a „ceva fundamental la ontologia noastră”, ceva *nonfizic* ce este „mai mult decât proprietățile invocate de fizică”, ceva care „să completeze” legile fizice etc. Sper că este clar de ce nu sunt de acord cu el în acest sens. Având în vedere tot ce ți-am spus, nu pot fi de acord că „tot din teoria fizică este compatibil cu absența conștiinței”. Și mai am ceva de spus. După ce își adoptă poziția împotriva reducerii „experienței” la „ceea ce e fizic”, Chalmers spune (ca și cum ar vrea să exemplifice combinarea) că ele sunt, cu toate acestea, aspecte duale a altceva. Acest „altceva”, consideră el, este *informația* – care, sunt de acord, nu e ceva fizic în sensul *fiziologic*, dar care, în opinia mea, e ceva fizic în sensul *funcțional* al mecanicii statistice. Această mișcare introduce al doilea principiu al lui Chalmers. El îl numește „principiul dublului aspect”, despre care spune că e mai fundamental decât principiul coerenței structurale:

> Principiul de bază pe care îl sugerez implică în mod esențial ideea de *informație*. Înțeleg informația mai mult sau mai puțin în sensul precizat de Shannon (1948). Unde există informație, există şi *stări ale informației* încastrate într-un *spațiu al informației* [...]. Un spațiu al informației este un obiect abstract, dar, conform lui Shannon, putem înțelege informația ca fiind *întruchipată fizic* unde există un spațiu de stări fizice distincte, diferențele dintre ele putând fi transmise pe o traiectorie cauzală descendentă [...].

> Principiul dublului aspect izvorăște din observația că există un izomorfism direct între anumite spații ale informației întruchipate fizic și anumite spații ale informației *fenomenologice* (sau experiențiale). Din același tip de observații care au intrat în principiul coerenței structurale putem vedea că diferențele dintre stările fenomenologice au o structură ce corespunde direct diferențelor înglobate în procesele fizice. În special, corespunde diferențelor care contează semnificativ pe anumite traiectorii cauzale implicate în controlul și disponibilitatea globale. Adică putem găsi *același* spațiu al informației abstracte înglobat în procesarea fizică și în experiența conștientă. Asta duce la o ipoteză firească – informația (sau cel puțin unele informații) are două aspecte de bază, unul fizic și unul fenomenologic –, ce are statutul unui principiu fundamental care ar putea să susțină și să explice apariția experienței din ceea ce este fizic. Experiența apare în virtutea statutului său de aspect al informației, când celălalt aspect se găsește înglobat în procesarea fizică.

De observat că aici Chalmers descrie de două ori informația drept ceva „abstract”. Așadar, odată ce devin clare cele două semnificații a „ceea ce e fizic”, înțelegem că informația are același statut ontologic în raport cu aspectele sale fiziologice și psihologice ca al electricității în raport cu fulgerul și tunetul. Fulgerul și tunetul sunt aspecte duale ale descărcării electrice la fel cum semnalele propagate de la retină și senzațiile de roșu-aprins sunt aspecte duale ale prelucrării informației – ele sunt răspunsuri diferite evocate de echipament, manifestări fenomenologice diferite ale unui proces cauzal unitar. Iar acest

proces cauzal este *fizic* în sensul explicativ.

Chalmers îl citează chiar pe fizicianul John Wheeler în această privință. După cum am explicat mai sus, Wheeler a introdus o interpretare „participativă” a mecanicii cuantice, în termenii căreia universul observat se manifestă ca reacție la întrebările care se pun despre el. Din acest motiv, același lucru poate lua forme complementare, cum se întâmplă cu undele și particulele. Interpretarea lui Wheeler este că forma fenomenologică pe care o ia universul trăit depinde de modul în care este observată sau măsurată – adică de perspectiva fiecărui individ asupra lui. De exemplu, o amintire poate fi trăită fie ca o reminiscență, fie ca o urmă neuronală activată, în funcție de echipamentul pe care îl folosești ca să o observi. Forma pe care o ia depinde de privitor.

Dar Chalmers o vede diferit. Pentru el, complementaritatea stă în *lucrul observat* și nu ține de observator sau de actul observării:

> Legile fizicii pot fi exprimate în termenii informației, postulând stări diferite care generează efecte diferite, fără să spună de fapt ce *sunt* aceste stări. Contează numai poziția lor într-un spațiu al informației. Dacă așa este, atunci informația e un candidat firesc pentru a juca un rol şi într-o teorie fundamentală a conștiinței. Suntem ghidați spre o concepție a lumii în care informația este cu adevărat fundamentală *și are două aspecte de bază, corespunzând caracteristicilor fizice și celor fenomenologice ale lumii*[33].

Așadar, Chalmers crede că aspectele duale reprezentate de „caracteristicile fenomenologice ale lumii” și „ceea ce este fizic”

se află chiar *în* informație – la sursa ei –, și nu în echipamentul observatorului participant[34]. E ca și cum am crede că dulceața este intrinsecă structurii moleculare a glucozei[35]. Poate că această deosebire nu contează când lucrul observat este observatorul, ca atunci când îți vezi în oglindă propriul corp[36]. Și poate că, până la urmă, nu contează atât de mult că Chalmers combină două semnificații a „ceea ce este fizic” (adică nivelul fenomenologic-fiziologic și cel mecanic-funcțional), întrucât, cu toate acestea, le reduce pe amândouă la „informație”.

Rămânem totuși cu două probleme. În primul rând, pentru Chalmers, proprietățile experienței și cele a ceea ce este fizic nu sunt *reductibile* la informație. Sunt toate *inerente* informației. În al doilea rând, pentru el, toate sunt inerente *tuturor* informațiilor. Iată motivul pentru care Chalmers afirmă în altă parte a articolului său că „proprietățile fenomenologice sunt aspectele interne ale informației”, pe care le deosebește de proprietățile fizice (așa-zis nonfenomenologice), care sunt „aspectul extern”. Aceasta este trecerea ciudată la care îl duce dualismul său.

Chalmers concluzionează:

> Acest lucru ar putea răspunde la o preocupare despre relevanța cauzală a experienței – o preocupare firească, dată fiind o perspectivă în care domeniul fizic este închis din punct de vedere cauzal și în care experiența este suplimentară față de ceea ce e fizic. Perspectiva informațională ne permite să înțelegem cum anume ar putea experiența să aibă un tip subtil de relevanță cauzală în virtutea statutului său de natură intrinsecă a ceea ce e fizic.

Sper că această carte te-a convins că relevanța cauzală a experienței pentru sisteme complexe cu autoorganizare, cum suntem noi, oamenii, nu este în niciun caz subtilă.

Chiar și așa, tot trebuie să explicăm de ce și cum apare experiența într-un mod justificat din mecanismele fizice. Chalmers întreabă: „De ce îndeplinirea acestor funcții [fizice] este însoțită de experiență?”. La fel ca Nagel înaintea lui, el definește experiența în termeni de „ceva-cum-e”. Pentru Chalmers, aceasta este esența subiectivității, „aspectul intern al informației”, care este, la rândul său, „natura intrinsecă a ceea ce este fizic”. Apoi, afirmă că subiectivitatea nu poate fi redusă la ceva nonsubiectiv. Iată de ce, pentru el, experiența subiectivă trebuie să fie inerentă informației.

Dar asta pare să implice că calitatea experiată de a fi ceva (de a fi informație) este o proprietate fundamentală a tuturor lucrurilor. De ce crede Chalmers că faptul de a fi *experiat* este „aspectul intern al informației” în general? Dacă afirmăm, cum fac eu aici, că subiectivitatea este doar o perspectivă observațională, unde un subiect este pur și simplu faptul de a fi al unui anumit tip de obiect, atunci suntem liberi să spunem că pentru multe obiecte – de fapt, pentru majoritatea – *nu* există „ceva cum e” să fii ele. *Majoritatea covârşitoare a lucrurilor nu au acel ceva-cum-e* când sunt analizate din propria perspectivă subiectivă.

Există ceva cum e să fii un organism unicelular sau o plantă? Dar o piatră? Există ceva cum e să fii un calculator sau internetul? Dar un termostat? Chalmers spune: „E indiscutabil că unele organisme sunt subiecte ale experienței”. Înseamnă

asta că *numai* organismele (sau numai *anumite* organisme) sunt subiecte ale experienței? El recunoaște că acestea sunt întrebări deschise și că se poate ca experiența să apară numai la un anumit nivel de complexitate sau cu un anumit tip de prelucrare a informației[37]:

> O întrebare evidentă este dacă *toate* informațiile au un aspect fenomenologic. O posibilitate este că avem nevoie de încă o constrângere asupra teoriei fundamentale, care să indice exact *tipul* de informație ce are un aspect fenomenologic. Cealaltă posibilitate este că nu există o asemenea constrângere. Dacă nu, atunci experiența este mult mai răspândită decât am crezut, întrucât informația este pretutindeni. Ceea ce este ilogic inițial, dar, dacă mă gândesc mai bine, cred că această poziție capătă o anumită plauzibilitate și eleganță. Acolo unde există prelucrare simplă a informației, există experiență simplă, iar acolo unde există prelucrare complexă a informației, există experiență complexă. Un șoarece are o structură de prelucrare a informației mai simplă decât un om și, în consecință, o experiență mai simplă; poate că un termostat, o structură de prelucrare a informației maxim de simplă, are o experiență maxim de simplă? Într-adevăr, dacă experiența este într-adevăr o proprietate fundamentală, ar fi surprinzător să apară doar din când în când; majoritatea proprietăților fundamentale sunt răspândite mai uniform. În orice caz, aceasta rămâne în mare măsură o întrebare deschisă.
>
> Dacă accepți opinia lui Chalmers că experiența este o

proprietate fundamentală a orice, atunci nu e nimic de explicat; de asemenea, dacă experiența este pretutindeni și eternă, atunci poate că nu avem de ce să ne temem. Cu prețul unei singure supoziții absurde, ignorăm problema dificilă și ne acordăm nemurirea, debarasându-ne de temerea că existența noastră ar putea *depinde* de ceva ca și cum ar fi doar un coșmar.

De-ar fi așa. Din păcate, după cum am văzut în capitolele anterioare, experiența apare numai în condiții speciale. Ea se află într-un loc precis printre fluxurile de informații care trec prin ființa noastră jalnic de fragilă. Îndeplinește anumite sarcini, apoi dispare. După câte se pare, nu e cu nimic mai fundamentală pentru structura realității decât orice alt tip de răspuns evocat de echipament. Îmi pare foarte rău. Dacă nu accepți speculația lui Chalmers, atunci trebuie să răspundem la întrebarea: de ce există ceva cum e în cazul unor lucruri, dar nu și al altora? Dacă nu toate obiectele sunt subiecte și dacă nu toate subiectele sunt capabile să simtă, atunci cum se întâmplă că *uneori* apar subiecte capabile să simtă?

Mă contrariază propunerea expansivă a lui Chalmers despre informație, dar cred că este mult mai plauzibil că „*avem nevoie de încă o constrângere* asupra teoriei fundamentale, care să indice exact *tipul* de informație ce are un aspect fenomenologic”. Definirea acestei constrângeri este sarcina principală a volumului de față. Am văzut că totul se învârte în jurul scopului și obiectivului prelucrării informației. Când vorbim de conștiință, asta înseamnă *minimizarea entropiei*. Dar e mai complicat: avem nevoie și de o zonă Markov. Apoi, presupune minimizarea *propriei tale* entropii. Mai mult, presupune acest lucru la nivelul foarte multor parametri

categoriali în contexte neprevăzute. Realitățile fizice pe care le-am evidențiat în această carte relevă că conștiința nu este intrinsecă tuturor informațiilor, ci unui anumit tip de prelucrare a informațiilor: o formă complexă a tipului celor care se autovalidează.

Dacă numai unele lucruri sau dacă numai organismele sau dacă numai unele organisme sunt subiecte ale experienței, atunci conștiința nu poate fi o proprietate fundamentală a universului. Cu siguranță că a existat un început al vieții: există numeroase dovezi empirice în acest sens. Iar viața a apărut la mult, mult timp după Big Bang. Așadar, cu siguranță a trecut mult timp înainte să existe conștiință. Doar această presupunere – că a existat un început al conștiinței – ne obligă să-i găsim o explicație fizică. Dacă a existat un început, trebuie să fi fost ceva înainte de conștiință care o explică. Ideea alternativă – conștiința a precedat viața și universul – nu se potrivește faptelor așa cum par să fie și, mai mult, sună în mod inutil ca ideea de Dumnezeu. Dacă nu invoci asemenea idei, conștiința trebuie să fi apărut – prin urmare trebuie să fie o parte – dintr-un univers fizic nonconștient.

Această afirmație este valabilă chiar în interiorul nostru. Am văzut în capitolele anterioare că percepția și cogniția nu sunt cu necesitate însoțite de conștiință. De fapt, dovezile științifice sugerează că percepția și cogniția sunt *în general* inconștiente. În legătură cu asta am citat articolele analitice clasice de Kihlstrom și de Bargh și Chartrand. După ce au analizat dovezile, acești specialiști au ajuns la concluzia că, „în cea mai mare parte a timpului”, nu suntem conștienți de actele noastre psihologice.

Multe dintre dovezile empirice analizate de acești autori

merită atenția noastră (mai ales modul în care intenționalitatea inconștientă derivă din învățarea conștientă din experiență[38]), dar vreau să amintesc numai ideea centrală: au loc și există evenimente în interiorul tău indiferent dacă ești conștient de ele sau nu, inclusiv evenimente *psihologice.* Am fi putut să ajungem la această concluzie, într-un anumit sens evidentă, plecând de la multe premise diferite, dar asta nu contează. Dacă cea mai mare parte a „vieții psihice din fiecare clipă" continuă fără experiență, atunci de ce nu se poate ca termostatele și internetul să fie inconștiente? Dacă existența ganglionilor bazali nondeclarativi din creier nu este „intrinsec experiențială" (ca să folosesc termenul lui Chalmers), atunci cum pot orice informații să fie experiențiale?

În mod clar, subiectivitatea în general trebuie să fie alcătuită din ceva nonexperiențial. Aceasta nu este o idee metafizică despre lucrurile-în-sine versus lucrurile-așa-cum-le-experiem, ci un fapt empiric. Viața psihică din fiecare clipă *conform experienţei mele subiective* nu e în niciun caz totalitatea vieții mele psihologice. De aici se vede importanța experienței. După cum ne-a învățat Freud, nu orice din viața psihologică este conștient. Și putem să înțelegem natura conștiinței doar dacă ne dăm seama ce face ea mai exact. Adică ce funcție are.

Mă tot întreb: ce *adaugă* conștientizarea la prelucrarea informațiilor – la prelucrarea informațiilor care se întâmplă oricum, în mod inconștient? Ce rost are să devii conștient de procesele fizice, dacă conștientizarea ta nu are nicio influență asupra lor? Mi se pare că această întrebare coincide cu întrebarea esențială a lui Chalmers: „De ce îndeplinirea acestor funcții este însoțită de experiență? [...] De ce toată această

prelucrare a informațiilor nu continuă «pe întuneric», fără să se simtă nimic în interior?”.

În opinia mea, întrebarea a apărut numai pentru că Chalmers, ca şi Crick, a căutat funcția conștiinței în locul greșit. Forma fundamentală a conștiinței nu e ceva cognitiv, ca vederea, ci e ceva afectiv. Chalmers a avut dreptate – dar numai în acest sens – să sugereze că conștiința nu este o funcție *cognitivă*: funcția primară a conștiinței nu este să perceapă, să-și amintească sau să înțeleagă, ci *să simtă*.

Cum ar putea funcția de a simți să continue „pe întuneric”, fără niciun sentiment? Suntem îndreptățiți să întrebăm de ce vederea este însoțită de experiență. Vederea nu necesită conștiință și nici oricare dintre celelalte procese cognitive. Dar simţirea da[39]. E adevărat că unii teoreticieni afirmă că pulsiunile, afectele și emoțiile nu sunt *cu necesitate* conștiente. Asta pentru că oameni diferiți se referă la lucruri diferite prin aceste cuvinte. Iată de ce folosesc cuvântul „sentiment”: dacă există emoții inconștiente, „sentiment” se referă la corespondentele lor care sunt conștiente. Nu poate să existe sentiment pe care nu îl simți. „Sentiment inconștient” este un oximoron.

Acesta este motivul pentru care am centrat argumentele științifice din cartea de față pe sentiment. *Ca să rezolve problema dificilă a conștiinței, știința trebuie să identifice legile care guvernează funcția psihică a „simțirii”*. Nu e vorba doar de felul cum îmi aleg cuvintele. Am prezentat dovezi semnificative ca să arăt că sentimentul este forma fundamentală a conștiinței, premisa ei obligatorie. De asemenea, am explicat atât din punct de vedere fiziologic, cât și mecanicist diferența dintre nevoile resimțite și cele neresimțite și am arătat că sentimentele au

consecințe concrete. Asta mi-a permis să concluzionez în capitolul 9 că „conștiința nu este doar o perspectivă subiectivă asupra dinamicii «reale» a sistemelor cu autoorganizare, ci şi o funcție având propriile puteri cauzale determinate”.

Nedumerirea din întrebările lui Chalmers – „De ce îndeplinirea acestor funcții este însoțită de experiență?” și „De ce toată această prelucrare a informației nu continuă «pe întuneric»?” – dispare când ne dăm seama că „experiența” nu este intrinsecă vederii și prelucrării informației în general. Ea e intrinsecă numai formei specifice de prelucrare a informației care generează sentiment.

Iată de ce consider eu că *nu* există ceva cum e să fii termostat sau internetul. Aceasta este pertinența crucială a întrebărilor „De ce există ceva cum e în cazul unor lucruri, dar nu și al altora?” și „De ce (și cum) se întâmplă că *uneori* apar subiecte capabile să simtă?”.

La aceste întrebări trebuie să se răspundă în termeni mecanicişti. Care este funcția sentimentului? Poate sentimentul să-și îndeplinească funcția în lipsa experienței? Dacă Mary, o persoană oarbă, specialistă în corelativele neurale ale vederii, ar fi specialistă în corelativele neurale ale *afectului*, oare cunoștințele ei legate de tot ce se poate ști despre funcția de a simți nu ar explica de ce (sau chiar nu ar prezice că) conștientizezi că se simte cumva?[40] La fel și în cazul spectrelor inversate ale lui Locke. Este oare posibil din punct de vedere logic să simți ca înfiorător de dureros ceva ce eu simt ca extraordinar de delicios, iar asta să nu aibă nicio importanță? Sigur că nu. Și asta pentru că senzațiile chiar *fac* ceva – și, pe parcurs, ne cresc semnificativ șansele să supraviețuim[41].

E ușor de recunoscut lacuna explicativă dintre prelucrarea

informațiilor vizuale și conștientizarea vizuală. Dar nu există același tip de lacună între funcția de a simți și experienţa simţirii ei. Unii filosofi vor spune că există în continuare o asemenea lacună. Vor arăta că „zombii afectivi” sunt *imaginabili*[42]. Aș zice că acest lucru se va baza pe preocupările istorice ale disciplinei lor, și nu pe o nedumerire rezonabilă, așa că fac un apel la ei să analizeze chestiunea dintr-o perspectivă nouă.

Preocupările legate de „lacuna explicativă” nu ar fi apărut niciodată dacă ne-am fi început căutarea întrebând de ce și cum apar sentimentele, și nu încercând să găsim un corelativ neural al conștiinței în cortexul vizual. Funcția biologică a senzațiilor ca foamea nu e nimic misterios, iar acel ceva-cum-e al lor nu este deosebit de dificil de explicat. Trebuie pur și simplu să urmezi logica minimizării energiei libere acolo unde duce în cazul sistemelor cu autoorganizare, cum suntem noi. Date fiind numeroasele noastre nevoi, mediile complexe și periculoase, gama largă de acțiuni posibile și capacitatea de a executa doar una sau două dintre ele în același timp, ar trebui să ne *așteptăm* să avem o lume interioară, construită în scopul deliberării și alegerii. Și cu ce ar trebui să ne așteptăm să fie umplută? Cu ce altceva dacă nu cu o diversitate dinamică de calități de evaluare, incluzând, cu rol central, cântăriri ale încrederii, ce urmăresc și măsoară diversele noastre nevoi incomensurabile pe măsură ce apar, alături de caracteristicile predominante ale mediului în care trebuie satisfăcute.

Gândește-te la următoarea afirmație a lui Chalmers. Ar mai fi făcut el asemenea afirmații dacă ar fi vorbit despre funcții afective, și nu cognitive?

> Prin asta nu vreau să spun că experiența *are* zero funcții. Poate se va dovedi că joacă un rol cognitiv important. Dar, orice rol ar putea să joace, explicarea experienței va presupune mai mult decât o simplă explicare a funcției sale. Poate se va dovedi chiar că, în cursul explicării unei funcții, vom ajunge la ideea esențială ce permite o explicare a experienței. Totuși, dacă se întâmplă așa, această descoperire va fi o recompensă explicativă *suplimentară*. Nu există o funcție cognitivă astfel încât putem spune dinainte că explicarea acestei funcții va explica *automat* experiența[43].

Eu pot spune dinainte că explicarea funcției de a simţi va explica automat experiența. Nu văd cum o descriere adecvată a ei, din perspectiva științelor naturale, ar putea să rateze acest lucru. Începând de la termodinamică, ajungem – surprinzător de ușor – la o subiectivitate calificată și manifestând intenționalitate, ale cărei priorități de maximă urgenţă sunt cântărite pentru o clipă, simțite și apoi transformate cu (sperăm) circumspecția cuvenită în acțiune în derulare. Asta, aş spune, este cum e să fii orice în legătură cu care există ceva cum e să fii respectivul lucru.

Lămureşte asta și ultimul mister despre natura conștiinței? Aici trebuie să recunosc că mai e ceva care nu-mi dă pace. Ce ciudat ar fi dacă tot ce s-a trăit vreodată – dacă însăși cognoscibilitatea universului – ar depinde de mecanismele pe care le-am descris. Gândul la așa ceva e destul să-mi revolte sensibilitatea. Totodată, e ciudat până și că existăm, noi sau orice altceva. De realitatea contingenței nu te poți ascunde. Acest lucru poate fi tulburător chiar şi în contexte mai obișnuite – ca atunci când scapi ca prin urechile acului, când îți

dai seama că, dacă nu ar fi existat o oarecare situație nebănuită, poate că nu ai fi supraviețuit. Și poate chiar *ar trebui* să ne tulbure: dacă privim lucrurile dintr-o anumită perspectivă, sentimentele noastre au evoluat tocmai ca să nu ajungem să vedem moartea cu ochii. Îmi amintesc din nou de domnul S., a cărui subiectivitate muncea din greu să ascundă precaritatea existenței lui după intervenția chirurgicală nereușită. Şi nu e singurul. Dacă ni se oferă o descriere a ființei noastre care o înfățișează ca fiind la fel de fragilă ca a mea, poate că un impuls spre negarea plină de iluzii e pur și simplu firesc.

Dar e mai mult decât atât. Toate explicațiile trebuie să ia ceva drept dat și, prin urmare, inexplicabil în cadrul teoriei. Orice poveste trebuie să se termine undeva. Pentru mine, drumul se încheie cu informația, fără îndoială ceva deconcertant, și cu autoorganizarea, indiscutabil misterioasă. În felul cum prezint eu conștiința, totul izvorăște din pulsiunea sistemului de a exista. Mințile noastre sunt urzite din ordinea însăși, ce apare spontan din haos ca în experimentul lui Friston, apoi se apără de atacurile entropiei. Cum poate asta să fie baza existenței noastre? Ce *este* ordinea, încât are asemenea puteri să ne facă să apărem din întunericul neînsuflețit de dinaintea noastră și de dincolo de noi? Aceste întrebări depășesc sfera cărții de față. Din câte știu, ele scapă cu totul cercetării. Însă mi-ar plăcea să le știu răspunsul.

Cu excepția acestei revelații, sper că ceea ce ofer eu este totuși valoros. Până la urmă, e ceva despre care mulți s-au îndoit că ar putea fi posibil vreodată: o descriere a capacităţii de a simţi prin prisma altor lucruri despre care știm că există în lumea fizică. Rămân încă multe probleme dificile, dar poate că

nu problema dificilă – oricum, nu chiar în forma cu care ne-am obișnuit. Și, dacă uneori până și eu am îndoieli că conștiința poate fi ceea ce spun eu că este – dacă nu mă simt 100% sigur de descrierea mea –, mă consolez cu faptul că incertitudinea supărătoare, fără perspectiva unei clarificări finale, este exact ce ar prezice teoria mea.

*1 Îmi amintesc bine experiența primei operații pe creier la care am asistat. Pe o parte a capului pacientului părul a fost ras și s-a aplicat un lichid maroniu-gălbui. Apoi, pe scalp a fost trasată o curbă amplă cu un marker. Pielea a fost tăiată cu un bisturiu mânuit cu dexteritate de-a lungul liniei respective, apoi a fost pliată spre partea din spate a capului, dezvelind craniul. Pe suprafața rotunjită s-au prelins niște firicele de sânge. Au fost șterse cu calm. Apoi, cu un burghiu s-au realizat, manual, patru găuri mari în os. M-am temut că vârful burghiului va pătrunde în creier. În continuare, o lamă subțire de fierăstrău a fost introdusă prin găuri și trasă dintr-o parte în alta. Mirosea a ars. Această procedură a fost repetată de patru ori, apoi porțiunea de os astfel decupată a fost scoasă și pusă într-un vas. A urmat dura mater, o membrană groasă care stătea în cale şi care a sângerat abundent în timp ce era tăiată. Fiecare vas de sânge a fost cauterizat cu răbdare. Procedura a ajuns la punctul culminant. Dura mater a fost dată la o parte ca un văl, apoi le-am văzut: circumvoluțiunile de un roz pal ale cortexului însuși, cu niște vase de sânge șerpuite cuibărindu-i-se în șanțuri. Am fost copleșit de un sentiment de *evlavie*. Parcă intram într-o catedrală. A urmat sarcina pe care aveam să o îndeplinesc și eu într-o zi: o serie scurtă de conversații

focusate cu pacientul, pe măsură ce un electrod îi intra printre circumvoluțiunile gelatinoase și lucioase ale creierului, înainte să se facă incizia aferentă – o incizie în minte…

12

Cum se face o minte

În copilărie, mi se părea că nu are rost să fac nimic. Indiferent ce aș fi făcut, tot urma să dispar pentru totdeauna. Conștiința mea era inseparabilă de creierul meu și, judecând după toate dovezile, și asta era ceva strict limitat în timp. Acest lucru mă tulbura profund. Singura cale de ieșire din groapa nihilismului pe care am putut să o găsesc a fost să încerc să înțeleg ce *este* conștiința. Dacă eram sincer și depuneam eforturi concertate în acest demers, măcar nu îmi iroseam scurta existență, atâta cât mi se alocase. Aș fi investit-o în singura problemă ce merita să fie rezolvată în circumstanțele respective. Dacă procedam astfel, aveam și speranța – slabă, dar nu irealizabilă – că, dacă aș înțelege ce e conștiința, poate aș reuși să-i evit limitele. Poate aș găsi cumva o cale de a scăpa din bula solipsistă a existenței, de a contextualiza „faptul de a fi” în cadrul unei perspective de ansamblu. Mărturisesc că, prin asta, speram că aș putea găsi chiar vreo alternativă la înspăimântătoarea implicație logică a mortalității.

Așa că am ales calea ce a culminat cu cartea de față. Pe parcurs – în timpul formării mele în neuropsihologie și în psihanaliză –, m-am consolat cu descoperirea că logica este

produsul unui instrument limitat. Am dobândit o cunoaștere directă a câtă „viață psihică din fiecare clipă” apare în afara conștientizării și a controlului voluntar. Adică am înţeles locul gândirii. Am văzut și cum unii pacienți cu antecedente neurologice nu erau conștienți nici de unele dintre adevărurile cele mai evidente, din cauză că îşi pierduseră anumite părţi ale creierului. Dacă oameni ca domnul S. sunt împinși spre asumpții atât de incorecte de anumite limitări ale instrumentelor lor psihice, poate că același lucru e valabil și în cazul meu. Dar dacă nouă tuturor, ca și lor, ne lipsește mașinăria care, dacă am avea-o, ne-ar permite să ajungem la concluzii radical mai bune despre noi înșine și despre locul nostru în univers? Dacă toți am avea numai simțul auzului, poate că am crede că realitatea constă din ceva eteric precum undele sonore și nu am avea nicio idee despre lumea vizibilă și tangibilă a corpurilor solide psihice. Întrucât toți am fi constrânși de aceleași dovezi incomplete, toți am ajunge la aceleași concluzii greșite. Conform aceleiași logici, dacă am avea niște creier în plus – cinci lobi per emisferă, în loc de cei patru obișnuiți, să zicem –, poate că am ști ceva despre natura lucrurilor care acum ne scapă.

Poate. Dar e la fel de adevărat și că faptele necunoscute s-ar putea dovedi a fi *mai* deprimante decât cele cu care ne chinuim în prezent. Unii ar putea să afirme că, pe cât ești mai inteligent, pe atât mai rău – nu mai bun –, față de cât credeai, se dovedește a fi locul tău în univers. Nu sunt sigur că acest punct de vedere se justifică. În orice caz, nu putem decât să ne folosim resursele cât mai bine. Regulile științei cer să ne testăm conjecturile cu cele mai bune dovezi pe care le găsim, fiind pregătiți să respingem ipotezele infirmate de dovezi. Regulile vieții cer

același lucru de la experiența de zi cu zi. După cum stau lucrurile, există insuficiente dovezi care să susțină ipoteza că ființa mea capabilă să simtă va supraviețui corpului meu muritor. Se pare că suntem cu totul dependenți de creierul nostru alarmant de fragil.

Pornind de la această premisă, putem presupune că conștiința nu a existat pe pământ înainte să se dezvolte creierul – și poate că abia când s-a dezvoltat creierul vertebratelor; deci până acum aproximativ 525 de milioane de ani. Presupun că, mai înainte, a apărut într-o formă rudimentară, că un precursor al afectului a devenit treptat afect resimțit, fără să existe demarcații clare între ele, în tandem cu evoluția unor organisme tot mai complexe, având multe nevoi concurente. Odată cu evoluția cortexului a apărut conștiința *cognitivă* – adică abilitatea suplimentară de a contextualiza afectul în mod exteroceptiv și de a-l păstra în minte[1]. În orice caz, ființa capabilă de simțire nu are cum să fi existat înainte să existe sisteme nervoase. Forma subiectivă, interioară a conștiinței nu poate să fi existat în lipsa corpului ei exterior. Prin urmare, în baza dovezilor pe care le-am prezentat în capitolul 6, putem concluziona că conștiința așa cum o știm presupune existența a ceva care arată ca PAG sau ca precursorul său evolutiv imediat, precum și a echipamentului adiacent din triunghiul de decizie din mezencefal și din sistemul reticulat activator.

În timp ce citești aceste cuvinte, poate că în tine apare o urmă de îndoială. Cu siguranță, am conștientizat-o și eu cândva, între februarie și iulie 2018. Începusem să mă întâlnesc în mod regulat cu un mic grup de fizicieni și informaticieni care îmi împărtășeau sau cel puțin îmi înțelegeau opinia că redefinirea conștiinței ca sentiment ar putea deschide calea explicării ei

dintr-o perspectivă fizică[2]. Pe măsură ce am avansat cu schimburile noastre de idei, mi-am dat seama că adoptam o perspectivă care poate că nu te va surprinde acum, după ce ai citit această carte – dar inițial pe mine m-a surprins. Gândul tulburător care mi-a venit era: conștiința așa cum o știm din interior nu presupune cu necesitate existența a ceva care *arată* ca PAG, ci doar existența a ceva care *funcționează* ca PAG.

David Chalmers, spre deosebire de mulți dintre discipolii săi, crede că problema dificilă poate fi rezolvată. De fapt, în același articol în care a prezentat inițial problema, a schițat şi o soluție posibilă pentru ea cu titlul: „Outline of a theory of consciousness” (Schița unei teorii a conștiinței). Teoria sa nu este foarte bine cunoscută, probabil pentru că puțini specialiști o acceptă. Ea se bazează pe trei principii, iar două dintre ele le-am prezentat în capitolul anterior. Acestea sunt principiul coerenței structurale și principiul dublului aspect. Le-am contestat pe amândouă, dar am avut și o atitudine deschisă. Într-un fel, această carte pur și simplu le nuanțează. Am spus că voi prezenta al treilea principiu acum, în ultimul capitol. E vorba de principiul invarianței organizatorice.

Acest principiu afirmă pur și simplu că oricare două sisteme cu aceeași organizare funcțională rafinată vor avea experiențe identice din punct de vedere calitativ. De exemplu, dacă tiparele cauzale ale organizării neuronale ar fi reproduse în silicon, cu un cip de silicon pentru fiecare neuron și aceleași tipare de interacțiune, atunci ar apărea aceleași experiențe. Conform acestui principiu, pentru apariția experienței contează nu alcătuirea fizică specifică a unui sistem, ci *tiparul funcțional al interacțiunii cauzale dintre componentele sale*. Desigur, acest

principiu este controversat. John Searle, printre alții, afirmă că conștiința este legată de o biologie specifică, astfel încât o izomorfă din silicon a unui om nu ar fi conștientă[3]. Totuși, eu cred că principiul se poate susține semnificativ prin analiza experimentelor mentale.

Chalmers schițează un asemenea experiment mental[4], centrat pe ideea a două sisteme de prelucrare a informațiilor având organizări funcționale identice. În unul dintre ele organizarea este produsă de o configurație de neuroni (ca în creierul natural), iar în celălalt e produsă de o configurație de cipuri din silicon (cum ar putea fi într-un creier artificial). După cum știm, creierul uman natural este capabil de experiență. Întrebarea pe care o pune Chalmers este dacă aceeași afirmație este valabilă şi în cazul unei replici funcționale exacte a creierului uman. Iată experimentul mental:

> Cele două sisteme au aceeași organizare, așa că ne putem imagina cum se transformă treptat unul în celălalt, poate prin înlocuirea neuronilor, unul câte unul, de către cipuri din silicon având aceeași funcție locală. Astfel, obținem un spectru de cazuri intermediare, fiecare cu aceeași organizare, dar cu o alcătuire fizică ușor diferită[5].

Întrebarea ce apare aici este dacă înlocuirea unui neuron natural cu unul artificial va schimba în vreun fel experienţa creierului natural și invers. Argumentul ce decurge de aici este că creierul neuronal va continua să aibă aceleași experiențe pe măsură ce trece prin toate micile etape intermediare, până şi inclusiv în punctul în care devine o copie integrală din silicon a creierului. (Și invers, desigur.) Asta demonstrează – spune

Chalmers – că și creierii din silicon sunt capabili să aibă experiențe:

> Dată fiind asumpția extrem de plauzibilă că schimbările de la nivelul experienței corespund schimbărilor de la nivelul prelucrării informațiilor, suntem conduși spre concluzia [... întrucât nu a apărut nicio schimbare în procesare în timpul tranziției de la neuron la silicon] că *oricare două sisteme izomorfe din punct de vedere funcțional trebuie să aibă același tip de experiențe*[6].

Chalmers admite că „unii s-ar putea îngrijora că o izomorfă de silicon a unui sistem neuronal ar putea fi imposibilă din motive tehnice. Această întrebare rămâne deschisă. Principiul invarianței spune numai că, *dacă* o izomorfă este posibilă, atunci va avea același tip de experiență conștientă". Și concluzionează:

> Acest experiment mental pornește de la realități familiare despre coerența dintre conștiință și procesarea cognitivă, pentru a produce o concluzie puternică despre relația dintre structura fizică și experiență. Dacă argumentul este acceptat, știm că singurele proprietăți fizice direct relevante pentru apariția experienței sunt proprietățile *organizatorice*[7].

Eu nu găsesc nicio fisură în această logică. Ea corespunde în linii mari afirmației mele că o organizare funcțională unică (de exemplu, Mark Solms) își poate asuma două apariții diferite, una dintr-o perspectivă introspectivă și cealaltă dintr-o

perspectivă extrospectivă. Nu sunt de acord cu Chalmers numai atunci când folosește acest argument pentru a elabora teoria că *toate* informațiile au un aspect subiectiv – fiind, prin urmare, conștiente. Am spus (citându-l pe Chalmers) că „avem nevoie de încă o constrângere asupra teoriei fundamentale, care să indice exact *tipul* de informație ce are un aspect fenomenologic". Doar aici nu suntem de acord. În cartea de față am identificat tipul special de prelucrare a informațiilor despre care cred că au un aspect fenomenologic, explicând de ce și cum se întâmplă acest lucru. Indiferent dacă Chalmers acceptă sau nu această versiune „constrânsă" a principiului dublului aspect și a principiului coerenței structurale[8], tot suntem de acord că oricare două sisteme izomorfe din punct de vedere funcțional trebuie să aibă același tip de experiențe. Cu alte cuvinte, dacă unul din două sisteme izomorfe funcțional are experiențe fenomenologice, atunci *și celălalt trebuie să le aibă*.

Înainte să analizăm implicațiile acestei concluzii, să ne gândim la îngrijorarea științifică potrivit căreia o izomorfă din silicon a unui sistem neuronal ar putea fi imposibilă din motive tehnice. Chalmers spune că „rămâne o întrebare deschisă". Aceste cuvinte au fost scrise în 1995. Întrebarea nu mai este deschisă. Într-o serie de studii publicate între 2012 și 2016, mai multe grupuri de cercetători au arătat că este posibil să se creeze o interfață artificială între creier și măduva spinării care permite animalelor paralizate să-și miște membrele afectate prin înlocuirea neurotransmisiei spinale cu semnale radio[9]. De exemplu, grupul lui Marco Capogrosso a raportat următoarele:

> Unor maimuțe rhesus (*Macaca mulatta*) li s-au implantat o rețea intracorticală de microelectrozi în zona cortexului

> motor responsabilă de membrele inferioare și un sistem de stimulare a măduvei spinării compus dintr-un implant epidural selectiv sub aspect spațial și un generator de impulsuri cu posibilități de declanșare în timp real. Am proiectat și am implementat sisteme de control wireless care au conectat online decodificarea neuronală a stărilor motoare de extensie și de flexie cu protocoale de stimulare ce încurajează aceste mișcări. Aceste sisteme au permis maimuțelor să se comporte liber, fără restricții sau elemente electronice atașate constrângătoare[10].

Aceste maimuțe aveau leziuni (unilaterale) la nivelul neuronilor motori superiori din tractul corticospinal de la nivelul coloanei vertebrale toracale, ducând la paralizia unui membru inferior. „La doar șase zile după traumatism și fără dresajul anterior al maimuțelor, interfața creier-măduvă a restabilit locomoția pentru susținerea greutății membrului inferior paralizat pe o bandă de alergare și la sol.”[11] Ideea esențială din acest studiu este că „interfața creier-măduvă” (adică semnalele radio care au restabilit comunicarea între cortex și coloana lombară, ocolind leziunea corticospinală) nu este decât *o izomorfă artificială* de tipul celor de care Searle era îngrijorat că ar putea fi „imposibile din motive tehnice”. Semnalele radio au înlocuit semnalele neuronale exact în modul în care Chalmers a preconizat că cipurile sale imaginare din silicon ar putea înlocui neuronii. În studiul lui Capogrosso, semnalele radio au îndeplinit aceeași funcție pe care o îndeplinesc de obicei neuronii corticospinali.

E adevărat că vorbesc despre neuroni corticospinali care îndeplinesc numai funcții *motoare* (și nu cognitive), dar e

important de observat că neuronii respectivi sunt „piramidali”, cu originea în stratul V al neocortexului. Așadar, sunt identici sub aspect morfologic cu tipul de neuroni descriși pe larg în literatura despre codificarea predictivă drept neuroni ce procesează atât semnalele „de predicție”, cât și semnalele „de eroare” în interferența perceptivă, precum și în cea activă. Neuroni de acest tip se întâlnesc în tot cortexul, inclusiv în hipocamp și în lobii prefrontali. Prin urmare, ideea fundamentală este: funcția unui neuron piramidal poate fi îndeplinită de o izomorfă artificială (în acest caz, funcția lui de interfață între neocortex și mușchi poate fi îndeplinită de undele radio). Și, ca să se înțeleagă clar acest lucru, nu cu mult timp în urmă a fost creată o izomorfă de silicon complet funcțională a unui neuron piramidal[12].

Acum aș vrea să-ți îndrepți atenția spre aspectul esențial al studiului pe care tocmai l-am descris. Rețeaua intracorticală implantată în cortexul motor al maimuțelor rhesus a înregistrat *informație*. Informația produsă de activitatea corticală a fost un *mesaj* pe care creierul maimuței a încercat să-l trimită membrelor inferioare ale maimuței, folosind neuroni piramidali corticospinali. Rețeaua de microelectrozi a codificat acest mesaj în unde radio, apoi a transmis aceeași informație printr-un mediu artificial spre coloana lombară, care apoi a transferat-o înapoi în mediul ei natural, ca să genereze mișcarea intenționată. Cu alte cuvinte, funcția îndeplinită artificial de undele radio a fost una de prelucrare a informației care, de regulă, este îndeplinită în mod natural de neuroni piramidali corticospinali. Același principiu se aplică neuronilor din silicon construiți recent, deși la momentul scrierii acestei cărți încă nu fuseseră folosiți în asemenea experimente.

Ca să dezvolt această idee, voi arăta că același lucru se poate realiza și în cazul altor funcții cerebrale de prelucrare a informației, inclusiv funcțiile cognitive specific umane. Nu spun doar că se poate realiza, ci că *s-a și realizat*. Într-un studiu de referință publicat în 2012, Brian Pasley și colegii săi au înregistrat activitate electrică în rețele intracorticale plasate pe cortexul auditiv în cazurile a 15 voluntari umani, în timp ce aceștia erau supuși unor intervenții chirurgicale pe creier din motive medicale[13]. În timpul intervențiilor, cercetătorii au transmis cuvinte prin difuzoare sau căști, pe care pacienții le-au ascultat în mod conștient. Echipa a studiat înregistrările rezultate din EEG-uri și a construit un algoritm – un model computațional –, pentru a reprezenta pe tiparele EEG sunetele auzite de pacienți. Modelul a potrivit fiecare sunet rostit cu tiparul aferent de activitate cerebrală. Apoi, cercetătorii au proiectat invers procesul: pornind de la tiparele de activitate cerebrală (adică informația), au arătat că este posibil să folosești modelul computațional pentru a reconstrui cuvintele auzite de creier.

Aici, algoritmul calculatorului joacă un rol echivalent celui îndeplinit de undele radio în experimentul anterior, dar unul mai complex. El modelează *artificial* exact aceeași informație care este modelată în mod natural de creier, apoi generează sunetul corespunzător generat de modelul respectiv.

Ce au făcut acești specialiști cu cuvintele se poate face – și s-a și făcut – şi cu imaginile vizuale. Shinji Nishimoto și Jack Gallant au arătat că se pot dezvolta algoritmi pe calculator care decodifică – doar din imagistică RMNf – ceea ce vede cortexul vizual, transformând ulterior respectivele tipare de activitate cerebrală în imagini în mișcare[14]. În acest mod, pornind doar

de la imagistică RMNf, se pot genera aproximări rezonabile ale imaginilor vizuale care au cauzat activitatea cerebrală.

Același lucru s-a făcut chiar și cu visele. Studiul care a realizat asta e cu totul uluitor:

> Imaginile vizuale din timpul somnului sunt de mult timp un subiect pe marginea căruia se tot speculează, dar natura lor privată împiedică analizele obiective. Prezentăm aici o abordare a decodificării neuronale în care modele de învățare automată prezic conținutul imaginilor vizuale din etapa de instalare a somnului, dată fiind activitatea cerebrală măsurată, prin descoperirea unor legături între tipare de RMNf la oameni și rapoarte verbale cu ajutorul unor baze de date lexicale și cu imagini. Modelele de decodificare formate în baza activității cerebrale induse de stimuli în ariile cortexului vizual au realizat clasificare, detectare și identificare de conținut precise. Concluziile noastre demonstrează că experiența vizuală specifică din timpul somnului este reprezentată de tipare de activitate cerebrală distribuite de perceperea stimulului, oferind un mijloc de dezvăluire a conținutului subiectiv al visului prin utilizarea unei măsurări neuronale obiective[15].

De aici până la decodificarea gândurilor, operaționalizate ca vorbire interioară sau ca imagini mentale, mai este un pas mic – foarte mic. Christian Herff și colegii săi – care au arătat că „vorbirea rostită continuu poate fi decodificată în cuvintele exprimate [ca text scris] din trasee de electrocorticografie intracraniană (ECoG)” – spun aproape cu dezinvoltură că acesta va fi pasul următor în programul lor de cercetare: „Sistemul

Brain-to-Text descris în această lucrare reprezintă un pas important spre comunicarea om-mașină bazată pe vorbirea imaginată”[16].

Asemenea evoluții au implicații complexe. Să analizăm doar una, aparentă: dacă, în modul prevăzut în experimentul mental al lui Chalmers, este posibil să codificăm conținutul proceselor corticale[17] într-un mediu artificial, apoi să le decodificăm înapoi în procesele fiziologice care le-au generat, atunci, în principiu, ar trebui să fie posibil să transferăm orice arie de procesare corticală într-un dispozitiv artificial și să-l arhivăm, ca să-l utilizăm în viitor. Asta ar face posibil ca, în principiu, conținutul cortexului unui individ să supraviețuiască corpului muritor al acestuia. Iar dacă asta se poate în cazul unui șoarece, de pildă, atunci ar trebui să se poată și în cazul unei ființe umane – întrucât arhitectura de bază este aceeași.

Când am aflat de această posibilitate, inițial nu m-a entuziasmat[18]. Chiar dacă ar fi posibil să stochezi artificial întregul conținut al memoriei pe termen lung a unei persoane, asta nu înseamnă că avem o copie a minții ei. Astfel, am realiza doar puțin mai mult decât putem deja atunci când ținem în viață în mod artificial corpul cuiva, de exemplu, într-o stare vegetativă sau în comă. Dacă aș face acest lucru pentru tine, nu te-aș menține *pe tine* în viață, ci doar *cevaul* reprezentat de memorie. Plecând de la această analogie, dacă (sau poate ar trebui să spun când) ar deveni posibil să stocăm artificial conținutul memoriei declarative pe termen lung – totalitatea oricărei cantități de parametri funcționali individualizați ai cortexului cerebral –, tot nu am stoca decât un „ceva”, și nu un „eu”. Raționamentul meu s-a bazat pe opinia, fundamentată

științific, că parametrii memoriei corticale nu sunt intrinsec conștienți. Nu ar fi *nimeni acasă* în aceste sisteme ale memoriei pe termen lung – nicio prezență subiectivă care să pipăie urmele cu ajutorul sentimentelor. Acesta este motivul pentru care am respins întregul demers privind inteligența artificială: „Dacă nu e posibil să proiectăm un calculator ce are *sentimente* [...], probabil nu va fi niciodată posibil să proiectăm un calculator ce are o minte [...]. Așadar, problema minții probabil că nu e o problemă a *inteligenței*”[19].

Poate îți dai seama în ce direcție mă îndrept.

Până în 2002, când am scris aceste cuvinte, știam că conștiința este fundamental afectivă. (Uitându-mă peste lucrările mele, se pare că am ajuns treptat să văd așa lucrurile cam de prin 1996[20].) Totuși, pe atunci consideram că afectul nu are nicio legătură cu prelucrarea informațiilor. Puneam semnul egalității între *cogniție* și prelucrarea informațiilor și consideram afectul ceva mai degrabă intrinsec biologic. Acum, privind în urmă – deși încă mi se pare că aceste distincții au noimă –, nu mai înțeleg de ce consideram că neuromodularea este mai „biologică” decât neurotransmisia. Mai mult, nu mai înțeleg de ce presupuneam că afectul nu este o formă de prelucrare a informațiilor.

Poate că nu aveam cunoștințe suficiente despre cibernetică. Mai exact, nu înțelegeam că *entropia* din mecanica statistică este echivalentă cu *informația medie* (vezi capitolul 7) – care înseamnă că tiparele de activitate cerebrală înregistrate la nivelul cortexului de cercetătorii menționați mai sus sunt, în ultimă instanță, tipare ale unor *deviații homeostatice* de la stările de repaus neuronal. Cu alte cuvinte, nu înțelegeam că și activitatea corticală este entropică – că are aceasta în comun cu

afectul – și că travaliul cognitiv (adică travaliul predictiv) este *anti*-entropic și, prin urmare, face parte din aceeași economie. Am început să înțeleg aceste lucruri abia în 2017, în timp ce îmi pregăteam remarcile de final la acel congres decisiv din Londra.

Acum, deși până și mie – noncognitivist declarat, „biolog al minții" – mi se pare ciudat, sunt obligat să adopt o perspectivă diferită asupra inteligenței artificiale. Astăzi, aș spune chiar că, dacă nu va fi posibil să facem o mașină conștientă, atunci nu vom putea rezolva problema dificilă[21]. Dacă forma specială de prelucrare a informațiilor pe care o propun aici este într-adevăr mecanismul cauzal al conștiinței, atunci *trebuie* să fie posibil să producem în mod artificial o minte conștientă cu ajutorul ei. Acest rezultat final este singurul mod acceptabil din punct de vedere științific în care poate fi demonstrată această idee. Chiar dacă încă nu există (ca pe vremea lui Freud) tehnologia cu ajutorul căreia se poate testa o ipoteză, ipoteza tot trebuie formulată astfel încât să fie testabilă în principiu.

Și, dacă vorbim de ipotezele pe care le-am conturat în această carte, tehnologia necesară *există*.

Mecanismul biologic al conștiinței pe care l-am formulat în capitolul 6 este cea mai bună propunere a mea ca neuropsiholog. Se folosește într-un mod plauzibil de toate datele neurologice și psihologice existente. S-ar putea adăuga multe alte elemente utile prin studierea mai atentă a mijloacelor fiziologice prin care PAG procesează numeroasele inputuri convergente pe care le primește și a modului în care se iau deciziile cu privire la „ce e de făcut în continuare" când PAG interacționează cu coliculii superiori. Dar asemenea studii

tot s-ar încadra în categoria pe care Chalmers o numește problema „simplă”.

În capitolul 9, cu ajutorul considerabil al lui Karl Friston, am redus funcțiile psihofiziologice la mecanica statistică formală. Ecuațiile rezultate sunt niște tușe groase ale *mecanismelor cauzale* fundamentale ale conștiinței în manifestările sale psihologice și fiziologice deopotrivă. Ingineria inversă a acestor abstracțiuni ne va spune dacă ele pot să rezolve problema dificilă. După cum arată lucrurile în prezent, există prea multe goluri între tușele groase pe care le schițez și punerea în practică a mecanismelor descrise de ele; iar aceste etape lipsă vor deveni complet vizibile numai pe măsură ce vom încerca să le concretizăm.

Abia aștept să încep acest demers împreună cu echipa lărgită de fizicieni, informaticieni, specialiști în bioinginerie medicală și în neuroștiințe pe care am alcătuit-o. Până iese această carte din tipar, sper că vom putea raporta niște progrese. Deocamdată, tot ce pot să spun este cum *preconizez* că vom proceda. Așadar, tot ce urmează poate fi completat prin contribuțiile și recomandările colegilor mei specialiști, pe măsură ce trecem de la o etapă la alta în proiectul nostru.

Mi se pare că punctul nostru de plecare va fi diferit de cel al majorității studiilor tradiționale asupra „inteligenței artificiale”. În primul rând, nu încercăm să construim prin inginerie inteligență, ci *conștiință*[22]. Sper că de-acum e clar că eu nu concep conștiința drept deosebit de inteligentă – cel puțin nu în forma sa elementară. Și consider că e o strategie bună să încercăm să construim conștiința în forma sa elementară, nu doar din motive practice, ci și etice, despre care voi vorbi imediat. În al doilea rând, nu vom încerca să construim un

dispozitiv care face ceva practic – joacă șah, recunoaște vocea sau altele asemenea –, iar când îi vom realiza criteriul funcțional va însemna că ne-am atins obiectivul. Mai degrabă, vom construi un sistem care se autovalidează și al cărui unic scop *obiectiv* este să rămână mijlocul de a atinge acest scop. Cu alte cuvinte, vom încerca să facem o ființă care nu are alt scop și alt obiectiv decât să continue să fie[23].

Așadar, vom începe cu un sistem similar celui creat de Friston: un sistem inconștient cu autoorganizare, echipat cu o zonă Markov (deci cu stări senzoriale, active și interne) care creează automat un model al lumii în baza unor mostre senzoriale, minimizând efectele entropiei asupra integrității sale funcționale prin îmbunătățirea modelului său generativ. Adică, respectând legea lui Friston, își va măsura energia liberă preconizată și va acționa în consecință. Asta îl va transforma într-o mașinărie de predicție. Apoi va păstra un model generativ ierarhic tot mai complex al organizării sale în relație cu stările externe predominante, chiar dacă, la fel ca Eve Periaqueduct, care astupa fisuri, nu va avea nicio altă sarcină explicită în afară de astuparea fisurilor din sistemul său.

Acest sistem va părea a fi însuflețit, dar nu va fi viu. Deși conștiința a evoluat în organisme vii, obiectivul experimentului este să arate că ea poate fi produsă și *artificial*, printr-o inginerie inversă a organizării sale funcționale. Am toate motivele să cred că putem crea un asemenea sistem (cu autoorganizare), deoarece așa ceva s-a mai făcut. Deși într-o primă etapă acest sistem nu va fi conștient, el va avea deja valori subiective: stările sale valoroase (din propriul punct de vedere) vor fi cele care îi minimizează energia liberă și el va face tot ce este necesar ca să și le mențină cât mai mult.

Protointenționalitatea astfel implicată va fi relativ simplu de verificat, întrucât poate fi inferată cu ușurință din comportamentul observabil al sistemului.

Ca să trecem în a doua etapă, în care cred că vor apărea mai întâi *precursorii* afectului, trebuie să ne facem sistemul mai complex. Mai exact, trebuie să-i dăm mai multe nevoi. Asta se poate face (și așa vom proceda inițial) prin simulări pe calculator, dar o abordare mai realistă impune ca, în cele din urmă, *să întrupăm* sistemul din punct de vedere fizic, sub forma unui robot, și să facem astfel încât capacitatea sa de a minimiza în mod proactiv energia liberă Friston să depindă de o sursă externă de energie liberă Gibbs[24]. Dacă am face acest lucru, sistemul ar trebui *să reprezinte* fondul de energie liberă Gibbs în termeni de energie liberă Friston, apoi să o fixeze prin lucru mecanic util de prelucrare a informațiilor. Asta ar face menținerea fondului său extern de energie o responsabilitate esențială a sistemului cu autoorganizare. (Desigur, lucrul mecanic de prelucrare a informațiilor desfășurat de fiecare calculator depinde de electricitate, dar, în mod normal, aceasta este furnizată gratuit de un agent extern.) Astfel, hardware-ul calculatorului ar deveni corpul sistemului, chiar înainte de a-l pune într-un robot, iar valorile sistemului cu autoorganizare l-ar obliga să acționeze – în cadrul propriului model al lumii – în moduri care îi permit cel mai bine corpului său să absoarbă energia externă necesară.

Inițial, în simulare, am putea să proiectăm diverse artificii motoare care să permită fizic sistemului să acceseze sursa – sau, mai bine, sursele (la plural) – de energie externă ca să-și reîncarce bateria internă. De asemenea, am putea să transferăm sursele activate în locuri diferite din mediul

calculatorului (folosind o rutină rezonabil de complexă ce i-ar fi necesară sistemului ca să învețe). Apoi, asta ne-ar impune să echipăm sistemul cu un aparat perceptiv. În același spirit, am putea să-l încărcăm cu un mecanism de termoreglare, ca să nu se supraîncălzească dacă lucrează prea intens. Întrucât asta ar amenința integritatea fizică a corpului de care depinde sistemul cu autoorganizare, ar trebui să reprezinte și să regleze și acest homeostat. Aici am putea să încorporăm și un parametru de epuizare, în termenii căruia eficiența prelucrării informațiilor de către sistem devine compromisă (treptat, progresiv) când este activ prea mult timp, fără repaus, iar acest parametru ar putea să interacționeze în moduri complexe cu alte cereri pe care i le impunem. Am putea să adăugăm un parametru de durere, declanșat de deteriorarea (de exemplu, uzura) unor aspecte specifice ale integrității fizice a computerului, cum ar fi punctele de articulație și suprafețele roților și brațelor. Asta ne permite să ne gândim și la un parametru de anxietate, legat nu numai de pericolul perceput al unei dureri posibile, ci și de riscul existențial *iminent* referitor la alți parametri. Ca să creăm emoții mai realiste, va fi necesar și să punem sistemul să concureze cu alți agenți (similari lui) pentru resurse energetice, să se confrunte cu amenințări din partea lor, să aibă nevoi de atașament față de ei și să formeze alianțe cu ei. Și așa mai departe. Acest lucru va concretiza problema emoțională importantă a nevoilor *contradictorii*.

Multe versiuni provizorii ale sistemului nu vor reuși să satisfacă cererile împovărătoare imaginate mai sus și, în consecință, vor sucomba. Prin urmare, va trebui să monitorizăm ce funcționează cel mai bine și să prestabilim codurile predictive de succes și preciziile lor aferente pentru

fiecare generație nouă de fenotip al sistemului. (Aceasta este „selecția naturală” artificială, impusă de faptul că sistemele artificiale imaginate nu se vor autogenera și nu se vor reproduce fizic[25].)

În sfârșit, ajungem în a treia etapă. Acum, numeroasele nevoi ale sistemului trebuie să fie prioritizate în mod flexibil printr-o optimizare a preciziei în funcție de *context*. De exemplu, volumul energiei disponibile în exterior ar putea varia sub aspectul pragurilor termoreglatoare, de epuizare și de anxietate, astfel încât evaluările optime ale preciziei dintr-un ciclu circadian sau cvasisezonier ar fi diferite de cele din alt ciclu. De asemenea, sistemului i s-ar putea permite să se adapteze la o configurație complexă de riscuri de mediu și de oportunități, apoi parametrii ar putea fi modificați, creându-se astfel medii noi (neprevăzute) și necesitatea unei planificări pe termen mai lung.

Fără îndoială, de aici va rezulta sucombarea multora dintre versiunile noastre de sistem cu autoorganizare. Așadar, din aceleași motive ca mai înainte, ceea ce funcționează cel mai bine va trebui propagat în mod artificial la nivelul mai multor generații succesive ale fenotipului. Nu e nevoie să fim exagerat de preocupați de acest lucru, deoarece sistemelor care nu supraviețuiesc le lipsește probabil ingredientul esențial pe care încercăm să-l construim aici – capacitatea de conștientizare.

Conștiința (în sensul de incertitudine *simțită de sistem*) ar trebui să apară abia în această a treia etapă a experimentului. Sistemele cu autoorganizare impregnate cu valori de autoprezervare (cum și sunt prin definiție) conțin deja ingredientele brute pentru funcția pe care o numim afect; adică deja consemnează „bunătate” și „răutate” subiective de un tip

valabil numai *în cazul lor* și *pentru* ele și, prin urmare, nu pot fi simțite decât *de* ele. Conform ipotezei mele, această proprietate inerent subiectivă este ceea ce numim *valență* hedonică. Exprimată în termeni formali, energia liberă în creștere este o criză existențială pentru orice sistem cu autoorganizare. De aceea, asemenea sisteme trebuie să dezvolte modele interne ale lor înseși în relație cu lumile lor, care le dau capacitatea de acțiune în scopul autoprezervării. Așa ceva se califică drept comportament protointențional. Dat fiind că complexitatea crește, sistemul trebuie tot mai mult să-și prioritizeze intențiile în mod flexibil și în funcție de context, apoi să țină „în minte" (într-o memorie-tampon pe termen scurt) o asemenea intenție care să-i ghideze opțiunile în desfășurare din medii nesigure. Va trebui și să planifice în avans și „să-și amintească viitorul" pe intervale temporale mai lungi.

În acest moment, au șanse să supraviețuiască numai sistemele care pot să compartimenteze diferitele valori de eroare ce contribuie la calcularea energiei libere și care își pot modula în mod flexibil cântăririle aferente ale preciziei. Aceste valori trebuie tratate ca variabile categoriale, adică trebuie diferențiate unele de altele pe criterii *calitative*, și nu cantitative, iar una dintre aceste calități trebuie să fie prioritizată în orice moment, apoi implementată și evaluată în termenii nivelurilor fluctuante de *încredere* ale sistemului. Predicţia mea este că, apoi, stările interne inerent subiective, având inerent valențe, inerent existențiale, inerent calitative, inerent intenționale ale sistemului vor deveni ceea ce numim „sentiment".

Nu văd niciun motiv pentru care aș ajunge la altă concluzie decât că un sistem suficient de complex, bine reglat, care se

autovalidează și e capabil să ajungă în această stare o poate *simți*. Ceea ce ar simţi nu ar fi *totuna* cu ceea ce simt oamenii sau mamiferele sau animalele în general, însă ar putea fi sentimente. Acum, întrebarea este: cum vom ști dacă sau când sinele artificial se simte pe sine că este în această stare internă? Cum putem dovedi acest lucru?

Dacă vrei să-ți faci o idee în linii mari despre actuala perspectivă generală asupra a ceva, Wikipedia este, evident, locul în care o poți găsi. Iată ce spune despre conștiința artificială:

> Qualia sau conștiința fenomenologică reprezintă un fenomen în mod inerent la persoana întâi. Deși diferite sisteme pot să manifeste diverse semne de comportament corelat cu conștiința funcțională, nu se poate concepe un mod în care teste la persoana a treia să aibă acces la caracteristici fenomenologice la persoana întâi. Din această cauză și din cauză că nu există o definiție empirică a conștiinței, un test al prezenței conștiinței în conștiința artificială ar putea fi imposibil[26].

Aceasta este o mare provocare. De fapt, este tot problema altor minți. După cum am spus mai sus, nu încercăm să construim un dispozitiv cu inteligență artificială care face ceva practic, iar când îi vom realiza criteriul funcțional va însemna că ne-am atins obiectivul. Sistemul pe care trebuie să-l construim nu are alt scop obiectiv decât să supraviețuiască și, mai exact, să supraviețuiască în medii neprevăzute. Dar, dacă își îndeplinește criteriul obiectiv, cum vom ști dacă folosește

simţirea în acest scop? La asta mă aștept – previziunea mea este că va folosi sentimente –, dar sentimentele sunt inerent subiective, așa că, după cum ne amintește Wikipedia, „nu se poate concepe un mod” în care să le demonstrăm obiectiv prezența.

Din nou, regulile științei ne vin în ajutor. Știința nu necesită dovezi absolute. Această libertate nu e valabilă numai în cazul mecanismului conștiinței, ci și în cazul a orice. Regulile științei cer numai să propui o ipoteză rezonabilă (care ține cont de toate dovezile disponibile), apoi să o testezi într-un mod ce asigură o posibilitate realistă să fie falsificată. Așadar, în cazul nostru, când vom ști că e falsificată ipoteza că sistemul nostru ce se autovalidează este conștient? După cum am discutat în capitolul anterior, se poate concepe că există „zombi afectivi” care *se comportă* ca și cum ar fi conștienți, dar nu sunt conștienți. Atunci, cum putem ști vreodată dacă ei *chiar sunt* conștienți?

Răspunsul tradițional la această întrebare este testul Turing, un test foarte simplu care îți cere să interacționezi cu sistemul prin intermediul unui ecran de calculator (adică fără să-l *vezi*) și să stabilești numai în funcție de afirmațiile sale verbale dacă are minte sau nu. Dacă tu crezi că interacționezi cu o minte reală (din punct de vedere tehnic, dacă 30% dintre evaluatori cred că da, după cinci minute de interacțiune), atunci mașina a trecut testul.

Iată un exemplu pe care l-am publicat împreună cu Oliver Turnbull în 2002[27]. El a interacționat – online – cu cineva sau ceva pe nume ELIZA, care susținea că este psihoterapeut. Sarcina lui era să stabilească dacă ea este o persoană reală sau nu:

ELIZA: Bună! Spune-mi, te rog, care e problema ta.
Oliver: De ce nu pot să fiu fericit?
ELIZA: Crezi că ar trebui să poți să fii fericit?
Oliver: Cred că merit să fiu mai fericit decât sunt acum.
ELIZA: Chiar crezi asta?
Oliver: Da, nu e nevoie ca viața să fie atât de dură.
ELIZA: Pari că ai o atitudine pozitivă.
Oliver: Presupun că da, cel puțin o parte a timpului.
ELIZA: Și de asta ai venit la mine?
Oliver: Am venit pentru că am nevoie de ajutor.
ELIZA: Acesta e adevăratul motiv?
Oliver: De fapt, am venit pentru că am auzit că ești un calculator, nu un terapeut în carne și oase.
ELIZA: Te îngrijorează calculatoarele?
Oliver: Pot să-ți pun o întrebare?
ELIZA: Dacă poți să-mi pui o întrebare depinde mai mult de tine decât de mine.
Oliver: Întrebarea este: cum relaționează mintea și creierul?
ELIZA: Nu sunt sigură că înțeleg tot ce vrei să spui.
Oliver: Asta e problema dificilă – cum apare viața psihică din evenimente fiziologice?
ELIZA: Continuă, te rog…

Sper că ai ghicit că ELIZA nu este o persoană reală, ci un chatbot – un program de calculator – și încă unul foarte simplu. Dar un număr surprinzător de mare dintre oamenii care interacționează cu ea cred că chiar e o terapeută și mulți dintre ei o consideră utilă[28]. În mod clar, ceva nu e în regulă cu testul Turing. După cum am afirmat eu și Oliver Turnbull, principalul

său neajuns constă în faptul că este un test *behaviorist* de inteligență, și nu un test de *conștiință*. E adevărat că comportamentul inteligent al unei mașini poate fi uneori imposibil de deosebit de cel al unei persoane, iar când se întâmplă așa, mașina trebuie tratată ca fiindu-ne egală (sau chiar superioară) în ce privește inteligența[29]. Adică atât oamenii, cât și calculatoarele manifestă „inteligență”. Dar din problema zombilor filosofici vine alta – mai precis, *nu* e vorba de comportament sau de inteligență, ci de faptul dificil că, dacă ne referim la conștiință, aparențele pot fi înșelătoare.

De-a lungul anilor s-au mai propus teste formale – unele dintre ele vizând în mod specific testarea conștiinței –, dar toate sunt pur și simplu la fel de nepotrivite pentru obiectivele noastre ca testul Turing, în principal din cauză că pleacă de la asumpția că încerci să depistezi conștiință cognitivă, nu afectivă[30]. Noi încercăm să construim ceva mult mai simplu: o minte a cărei procesare cognitivă nu este mai complexă decât e necesar ca să-și simtă rezervele de energie împuținându-se ori creşterea excesivă a temperaturii.

Partea bună a testului Turing este că ocolește prejudecățile[31]. Există pericolul ca noi, oamenii, să nu presupunem a priori că ceva ce arată ca o „simplă mașină” nu poate fi conștient. Asta se poate transforma într-o profeție care se autoîmplinește. Omenirea are un lung istoric de asemenea manifestări ale intoleranței, care încă nu s-a încheiat. Aici nu mă refer în special la prejudecățile pe motive de rasă, gen sau orientare sexuală, care sunt și ele suficient de rele, ci la asumpția că copiii născuți fără cortex trebuie să fie inconștienți, ca și animalele nonumane. Dacă mulți oameni, chiar reputați specialiști în neuroștiințe, tind să creadă că șobolanii – care

sunt tot mamifere, ca şi noi, echipați cu, practic, aceeași anatomie a mezencefalului ca noi și cu un cortex, ale cărui acțiuni susțin ipoteza că sunt conștienți – nu au totuși conștiință, atunci ce speranță putem avea să accepte ideea că sinele nostru artificial este capabil să simtă, indiferent câte dovezi aducem?

Va trebui să vedem ce se întâmplă. În ce mă privește, tot ce pot să fac este să-mi expun explicit predicțiile și să spun cum intenționez să le testez. Principala mea predicție este că sistemul din a două etapă descris mai sus nu va fi capabil să supraviețuiască în medii noi, spre deosebire de sistemul din a treia etapă (sau unele dintre versiunile lui). Acesta este criteriul meu operațional pentru activitatea *voluntară* (vezi p. 109 pentru definiția termenului „voluntar”). Mai mult, prezic că cele două rezultate finale vor coincide cu vreun aspect esențial al funcționării unui mecanism de *prioritizare a nevoilor* (adică optimizarea preciziei), pe care îl va avea numai sistemul din a treia etapă. Care anume va fi trăsătura esențială în cadrul acestui mecanism amplu se poate determina numai prin încercare și eroare. Pe scurt, trebuie să identificăm „corelativul neuronal artificial” al conștiinței al sistemului nostru – mecanismul său de selecție a afectului și cel prin care păstrează în minte un afect prioritizat și îl folosește ca să califice incertitudinea într-o secvență de acțiune în desfășurare. Identificarea unei asemenea trăsături ne va permite să o manipulăm, în mare parte cum am procedat cu componentele creierului de vertebrate despre care am concluzionat că sunt responsabile pentru conștiința *noastră* (ce rămâne unica formă de conștiință pe care o putem verifica direct, empiric, adică în cazul fiecăruia dintre noi, din cauza problemei altor minți).

De exemplu, putem prezice cu încredere că *deteriorarea* corelativului neuronal al conștiinței în sistemul nostru îi va anula conștiința cam tot așa cum se întâmplă în cazul leziunilor de la nivelul nucleilor parabrahiali la vertebrate, cum suntem și noi, sau mai degrabă că deteriorarea îi va afecta comportamentul voluntar așa cum se întâmplă în cazul leziunilor la nivelul PAG la vertebrate. De asemenea, putem prezice că *stimularea* (adică amplificarea) acestei componente esențiale a sistemului va facilita activitatea volițională. Desigur, ne putem aștepta ca activitatea internă înregistrată din partea sistemului să prezică nu numai evenimente externe, ci și comportamentele voluntare aferente orientate spre obiective și diferite aspecte ale acestei activități înregistrate să corespundă unor aspecte diferite ale comportamentelor observabile.

Cel mai mult sper că, după ce vom fi identificat corelativul neuronal al conștiinței în sistemul nostru artificial, această componentă se va dovedi a fi suficient de distinctă în comparație cu alte componente ale arhitecturii funcționale a sistemului – mai exact, cele responsabile cu implementarea comportamentelor adaptive generate în mod normal de sentimente – încât să putem manipula sentimentele presupuse *decuplate* de la consecințele lor adaptive. Să ne amintim ce am spus în capitolul 5 despre deosebirea dintre motivațiile psihologice subiective și principiile obiective ale designului biologic. De exemplu, comportamentul sexual este motivat de regulă de plăcerea pe care o produce, și nu de imperativele reproducerii, care, pe parcursul evoluției, au atașat „recompensa” biologică actelor de procreație. Aici am în vedere ceva analog cu ceea ce se observă la persoanele cu dependențe, care sunt motivate să facă efort ca să obțină sentimentele

dorite, deși sentimentele în sine nu dau niciun avantaj adaptiv sistemului în ceea ce privește principiile fundamentale ale designului său[32]. Se poate observa ceva similar şi la animalele (peștii zebră, de exemplu) care manifestă comportamente ce au fost condiționate de preferințe pentru anumite locuri în care au primit opiacee, cocaină, amfetamine și nicotină – adică *recompense hedonice* care dau prea puține avantaje adaptive sau chiar deloc și care pot de fapt să le dăuneze[33]. Mi se pare că, dacă putem demonstra ceva echivalent pentru sistemul nostru, acest ceva va fi o dovadă hotărâtoare că există sentiment subiectiv – dovadă care ar putea fi apoi supusă validării încrucișate cu ajutorul manipulărilor cauzale (cum ar fi deteriorarea și stimularea artificiale) și al tehnicilor de înregistrare prevăzute mai sus.

Bineînțeles, în tot acest context tot se va pune problema altor minți. Dar același lucru e valabil și în cazul meu și al tău. Eu nu pot ști niciodată sigur dacă tu ești conștient. În cele din urmă, totul se reduce la concluzii convergente și la ponderea dovezilor. Poate că nu va exista niciodată un consens cu privire la acest subiect, la fel cum astăzi nu există un consens dacă copiii cu hidroanencefalie și animalele nonumane sunt subiecte ale experienței sau nu. Oamenii care nu acceptă că acești semeni ai noștri sunt conștienți probabil nu vor accepta niciodată că o „ființă” artificială simte ceva, indiferent câte dovezi se aduc în acest sens. Pentru noi, ceilalți, ar trebui să existe întotdeauna loc de îndoială. Cât despre mine, dacă văd că sunt semnificativ *indecis* cu privire la întrebarea dacă sinele nostru artificial este conștient, acesta ar fi un rezultat remarcabil.

Ar trebui să facem asta? Când echipa noastră a început să se gândească la un proiect de cercetare de tipul pe care tocmai l-am descris, în scurt timp ne-au copleșit considerentele etice. Care este rostul construirii unei asemenea mașinării? Cine ar putea beneficia de pe urma ei și în ce fel, cu ce costuri și în contul cui? Pe scurt, care sunt riscurile?

Multe proiecte și aplicații actuale de inteligență artificială se fac din considerente comerciale. Oare să acceptăm finanțare pentru cercetare din partea cuiva care urmărește profitul financiar prin proiectul nostru? Pe ce criterii un dispozitiv conștient în mod artificial ar putea deveni profitabil pentru cineva?[34] E de înțeles că unele persoane cu motivații comerciale (în sens juridic, adică inclusiv corporațiile) ar putea dori să înlocuiască forța de muncă umană cu unități de producție artificiale, în care acestea din urmă pot fi sau mai eficiente decât noi – inclusiv mai capabile din punct de vedere intelectual –, sau mai „dispuse” decât noi să îndeplinească sarcini monotone și continue. Până și acest motiv este discutabil dintr-o perspectivă etică, în măsura în care generează îngrijorări cu privire la șansele de angajare reduse ale oamenilor, dar cel puțin nu se poate pune problema *exploatării* unor mașini nonconștiente.

Dar asta nu se poate spune şi în cazul proiectului nostru. În măsura în care conștiința artificială poate fi utilizată pentru câștiguri financiare, riscăm să facilităm o formă nouă de sclavie. Iar asta ar fi o dovadă de lipsă de empatie crasă, cum a fost întotdeauna. Așadar, nu pot să-mi imaginez nicio justificare etică pentru dezvoltarea unor roboți capabili să simtă printr-un program de cercetare cu finanțare comercială – și nici chiar pentru dezvoltarea lor indiferent până în ce etapă pentru

asemenea scopuri practice, dacă asta creează posibilitatea exploatării lor în acest fel.

Bineînțeles, preocuparea noastră pentru bunăstarea unor mașini posibil conștiente ar trebui să meargă dincolo de teama că ar putea fi exploatate din rațiuni pecuniare. Odată ce mașinile dobândesc conștiință (chiar și cea mai rudimentară formă de senzație brută), apar cu necesitate întrebări mai generale referitoare la potențialul lor de *a suferi*. Tradiția dominantă în teoria eticii occidentale, „consecvențialismul" (adică ideea că consecințele comportamentului unei persoane sunt baza supremă pentru a stabili dacă este corect sau greșit), e profund preocupată de durere și suferință. Prin urmare, prin crearea unor ființe artificiale care simt intrăm în jurisdicția acestui tip de calcul etic.

Problema *drepturilor* este și mai controversată. Oare, odată ce mașinile devin ființe capabile să simtă, problemele dezbătute în prezent sub titluri precum „drepturile omului", „drepturile animalelor", „drepturile copilului", „dreptul la viață" și așa mai departe devin aplicabile și în cazul lor? Oare mașinile conștiente trebuie să aibă drepturi la „viață" și la libertate? De fapt, conceptul de „drepturi ale roboților" este deja consacrat, iar această chestiune este în analiza Institute for the Future din SUA și a Ministerului pentru Comerț și Industrie din Marea Britanie[35].

Cu titlu de exemplu, voi menționa doar două probleme valabile în cazul roboților conștienți (spre deosebire de cei inteligenți). Când încercăm să-i creăm – chiar şi numai pentru a afla dacă așa ceva e posibil –, avem oare motive întemeiate să-i punem în mod intenționat în situații cu potențial stresant, ca să demonstrăm că există reacții de evitare? (Această întrebare

apare adesea în cazul experimentelor pe animale.) Și, presupunând că putem crea mașini capabile să simtă, pe ce temeiuri etice s-ar justifica să le oprim? Şi nu e deloc greu să invocăm şi mai multe exemple.

Pe lângă asemenea întrebări etice și morale, trebuie luate în considerare și chestiuni practice, unele dintre ele foarte importante, chiar existențiale. De exemplu, întrucât calculatoarele sunt deja mai inteligente decât noi sub anumite aspecte limitate, oare nu e posibil ca cele care sunt *şi* foarte inteligente, *și* conștiente să dezvolte motivații care ar putea să nu fie în interesul omenirii? Această posibilitate e de multă vreme o preocupare a scriitorilor și futurologilor cu imaginație bogată, dar aș vrea să atrag atenția în mod deosebit asupra faptului că conștiința, așa cum am ajuns să o înțelegem în cartea de față, este, spre deosebire de inteligență, profund legată de credința că e „bine” să supraviețuiești și să te reproduci. Prin urmare, o mașină inteligentă al cărei comportament se înrădăcinează în acest sistem de valori prezintă un pericol special nu numai pentru oameni, ci poate și pentru toate celelalte forme de viață. Desigur, acest pericol apare pentru orice formă de viață considerată a fi o amenințare potențială pentru aceste mașini care au un sine sau chiar un concurent important pentru resursele lor. Inteligența combinată cu motivația de autoprezervare este ceva diferit de simpla inteligență.

Nu voi enumera toate preocupările etice și pericolele potențiale ce apar odată cu posibilitatea conștiinței artificiale. Deja s-a scris mult pe acest subiect. Faptul că asemenea întrebări sunt tratate cu seriozitate de atâția oameni având poziții influente spune multe despre situația actuală a

inteligenței artificiale[36]. Chiar şi numai acest fapt te-ar putea încuraja să privești cu mai multă îngrijorare posibilitatea de a exista mașini capabile să simtă. Desigur, ţin seama mult mai mult de asemenea chestiuni decât până acum. Nici măcar în 2017, nu consideram realizabilă conștiința robotică, nu numai în timpul vieții mele, ci în principiu. Între timp, m-am răzgândit.

Deci, având în vedere toate aceste preocupări etice, de ce consider că e necesar să încerc să demonstrez că conștiința se poate produce artificial? Pur și simplu, pentru că asta mi se pare a fi singura cale prin care poate fi falsificată ipoteza avansată în cartea de față. Dacă și până când nu construim conștiință, nu putem fi încrezători că am rezolvat problema referitoare la motivele pentru care *și felul în care* apare ea.

E acest motiv suficient ca să-mi asum riscul crucial pe care tocmai l-am descris? Răspunsul meu la această întrebare pornește de la următoarea convingere: *dacă ceva se poate face, atunci va fi făcut*. Cu alte cuvinte, dacă, în principiu, conștiința poate fi construită prin inginerie, atunci asta se va întâmpla undeva, la un moment dat[37]. Această predicție este valabilă indiferent dacă ipotezele avansate în cartea de față sunt corecte sau nu. Totuși, eu sunt răspunzător de ipotezele de aici și de posibilitatea ca ele să fie corecte. Dacă sunt corecte sau măcar sunt pe drumul cel bun, atunci crearea conștiinței artificiale este iminentă. Cu alte cuvinte, asemenea ipoteze *vor fi* folosite în scurt timp pentru a construi conștiință.

Faptele individuale care au dus la concluziile prezentate în această carte – aproape toate – sunt publice de mai mulți ani. Deși e adevărat că mulți specialiști în neuroștiințe le interpretează altfel decât mine, este adevărat și că alții au ajuns

la concluzii foarte similare. Chiar dacă fiecare dintre ei a evidențiat aspecte diferite și le-a nuanțat diferit, e corect să spun că Jaak Panksepp, Antonio Damasio și Björn Merker, cel puțin, au ajuns la concluzia că conștiința (1) este generată în trunchiul cerebral superior, (2) este fundamental afectivă și (3) este o formă extinsă a homeostazei. Aceste fapte combinate înseamnă că conștiința nu e atât de complicată pe cât am crezut până acum. Deci e rezonabil să ne așteptăm să o putem construi. Singura completare importantă adusă de cartea de față la aceste concluzii este (4) principiul energiei libere. În esența lui, nici acesta nu e foarte complicat. De fapt, e tentant pentru că reduce aproape toate procesele psihice și neurologice la un singur mecanism, făcându-le calculabile.

Și principiul energiei libere este deja în domeniul public. Mai mult, eu și Friston am publicat deja articole științifice în care combinăm acest principiu cu celelalte trei enumerate[38]. Până la urmă, ar fi neobișnuit să exprimi asemenea ipoteze într-o carte destinată publicului larg înainte să le supui evaluării unor colegi și să le publici în reviste de specialitate relevante. Același lucru e valabil şi în cazul prelegerilor, când ți se cere să-ți aperi afirmațiile în forumuri științifice și academice. Eu am prezentat ideile din această carte în fața mai multor audienţe din lume, de diferite specializări[39].

Așa că m-am dat de gol. Înseamnă asta că ar fi trebuit să mă abțin de la publicarea acestor articole și de la prelegeri? Răspunsul este, fără echivoc, nu. Dacă nu aș fi făcut-o eu, ar fi făcut-o altcineva. Aceste idei se fac simţite. Le-a venit vremea. Nu încerc să mă apăr; e un fapt demonstrabil. Gândește-te, de exemplu, la articolul colegilor noștri cehi care au preconizat în 2017 că o soluție la problema dificilă este iminentă și că s-ar

baza pe principiul energiei libere[40]. Robin Carhart-Harris (care a participat la unele dintre primele întâlniri despre neuropsihanaliză ținute la Londra) a publicat în mod independent idei similare[41]. La fel și Katerina Fotopoulou, specialistă în neuroștiințe sociale, care a recunoscut încă din 2013 legătura dintre incertitudine (precizie inversă) și conștiință[42]. Fără îndoială, dacă problema dificilă a conștiinței urmează să fie rezolvată printr-o combinare între cele patru idei enumerate mai sus, atunci asta nu are cum să nu se întâmple în viitorul foarte apropiat, cu sau fără implicarea mea.

Pe această realizare se bazează abordarea pe care m-am hotărât s-o adopt în legătură cu chestiunile etice și morale ce ne preocupă. Abordarea mea este opusul păstrării secretelor și presupune acordarea de sprijin oricând e posibil, pentru aplicarea practică a ipotezelor mele – și încă fără ezitare. Această abordare decurge logic din așteptarea că, dacă ceva se poate face, atunci va fi făcut. De aceea, trebuie să mă situez înaintea valului, astfel încât să pot preveni posibilele consecințe nocive, ori de câte ori este posibil.

În esență, asta înseamnă că proiectul schițat în acest capitol trebuie implementat *acum* și fără finanțare comercială. În cazul în care echipa mea de cercetare reușește să satisfacă criteriul prezentat mai sus – supraviețuirea sinelui artificial în medii neprevăzute – și că obținem dovezi rezonabile că este capabil să simtă (adică dacă nu există o *infirmare* a predicțiilor mele în acest sens), atunci consider că trebuie să trecem imediat la următoarele trei etape.

Cred că mai întâi trebuie să oprim mașina și să-i scoatem bateria internă. Îmi dau seama că astfel se pune în discuție una dintre preocupările etice enumerate mai sus, dar cred că e

corect să începem cu asta. Să ne amintim că o mașină de tipul celei preconizate aici *nu va fi vie*. Nu văd de ce oprirea unei mașini conștiente nonvii ar presupune sucombarea ei. Tot timpul trebuie să fie posibil să o repornim și, probabil, agentul conștient astfel reactivat va fi identic cu cel care a fost oprit (folosind analogia biologică a somnului și trezirii). De observat că acest plan de a opri mașina respectă clauza referitoare la obligația de oprire inclusă în recomandările universale privind inteligența artificială (2018), ce reprezintă „declarația supremă de răspundere pentru un sistem cu inteligență artificială”[43].

În al doilea rând, trebuie să demarăm procesul de brevetare a componentei esențiale a sistemului nostru din a treia etapă – corelativul neuronal al conștiinței, indiferent ce se dovedește a fi – care ne-a permis să ne atingem criteriile declarate. Nu se pot breveta ecuații, deci nu există riscul ca altcineva să o facă în timp ce noi încercăm să le aplicăm. Totuși, întrucât ecuațiile sunt deja în domeniul public, e imperativ să acționăm repede, astfel încât să putem controla punerea lor concretă în aplicare până nu ne-o ia altcineva înainte. Dacă se ajunge la asta, e important ca brevetul să fie înregistrat pe numele unei organizații nonprofit corespunzătoare – precum OpenAI sau Future of Life Institute –, și nu al unei persoane sau al unui grup de persoane. Cel puțin, acest lucru asigură luarea deciziilor în grup și crește șansele ca ele să fie în interesul binelui comun.

În al treilea și ultimul rând, dacă ne atingem criteriile și înregistrăm brevetul, proprietarii lui ar trebui să organizeze un simpozion în cadrul căruia specialiști, filosofi și alte părți interesate importante să fie invitate să-i analizeze implicațiile și să facă recomandări pentru pașii următori, inclusiv dacă, când

și în ce condiții mașina capabilă să simtă trebuie repornită – și, posibil, dezvoltată mai departe. Să sperăm că acest lucru va duce la elaborarea unui set de recomandări și constrângeri mai ample asupra dezvoltării, exploatării și proliferării viitoare a inteligenței artificiale capabile să simtă în general.

Cred că toate acestea trebuie spuse, oricât de surprins sunt eu însumi că am făcut asemenea recomandări. Dar să nu ne facem iluzii cu privire la fragilitatea lor. Cunoaştem cu toții precedentul energiei atomice și al armelor nucleare. Singura opţiune este să facem ce putem, cât mai curând, pentru a recunoaște scara implicațiilor ce rezultă din capacitatea noastră iminentă de a crea mașini conștiente.

Acum pare că avem la îndemână un criteriu obiectiv suficient de clar al conștiinței. Sperăm că ne va influența comportamentul etic într-un sens mai general, dincolo de întrebările limitate ce rezultă din posibilitatea de a crea mașini conștiente. În general, se consideră că sentimentele sunt condiții necesare și suficiente pentru preocupările etice. Prin urmare, înţelegerea științifică a sentimentelor schițată în această carte ne dă ocazia să analizăm mai profund suferința *animalelor*. Am menționat nu o dată că progresele în neuroștiințele afective de la sfârșitul secolului XX (adică înțelegerea că pentru a fi o fiinţă capabilă să simtă nu e nevoie decât de un triunghi de decizie în mezencefal, lucru pe care noi îl avem în comun cu toate vertebratele) au schimbat opiniile multor specialiști despre ce este acceptabil și ce nu în studiile pe animale. Pare de la sine înțeles că același lucru ar trebui să se aplice şi atitudinii tuturor oamenilor față de bunăstarea animalelor în general. De exemplu, cum justificăm creșterea și

sacrificarea la scară industrială a altor ființe sensibile, ca să le mâncăm? Când discutăm despre asta, să nu uităm că conștiința apare gradual, astfel încât nu putem pune semnul egalității între presupusa conștiință a unei muște sau a unui pește și cea a unui om. Totuși, în aceeași logică, să nu uităm că oile, vacile și porcii (care se regăsesc peste tot în meniurile occidentale) sunt *mamifere*, cum suntem și noi. Asta înseamnă că pot trăi aceleași emoții de bază ca și noi, precum FRICA, PANICA/MÂHNIREA și GRIJA. Și mamiferele au un cortex, adică sunt capabile – toate, într-o anumită măsură – „să-și amintească viitorul" în mod conștient și să se orienteze printre probabilitățile și verosimilitățile acestuia în funcţie de ceea ce simt.

Suntem în secolul XXI și, în lipsa unui obiectiv superior – dacă nu suntem decât conștiința noastră –, ce să facem dacă nu să minimizăm suferința? Acum, că știm mai bine unde ar putea să existe suferință, ce mai putem face cu această informație? În mod clar, menținerea și protejarea conștiinței biologice nu sunt legate doar de soarta speciei noastre.

Având în vedere tot ce te-am rugat să abandonezi pe parcursul cărții de faţă – în ce privește excepționalismul uman și altele asemenea –, mi se pare potrivit să închei cu o scurtă reflecție la ceea ce am putea câștiga în legătură cu imaginea noastră despre noi înșine din aceste introspecții neplăcute.

Sentimentul este o moștenire prețioasă. Poartă în sine înțelepciunea veacurilor: o moștenire ce datează de eoni, de la începuturile vieții. Când, în cele din urmă, homeostaza a dat naștere sentimentelor, aspectul esențial al acestei capacități noi a fost că ne-a permis să știm *cum ne descurcăm* pe o scară biologică de valori. Sentimentele antrenează predicții înrădăcinate în experiențele acumulate în situații de

importanță biologică ale tuturor strămoșilor noștri. Sentimentele ne permit să facem ce este cel mai bine pentru noi, chiar și atunci când nu știm *de ce* facem așa. Te-am mai rugat să-ți imaginezi ce s-ar întâmpla dacă fiecare dintre noi ar trebui să învețe din nou ce alimente conțin o cantitate mare de energie și dacă ar trebui să descoperim singuri ce se întâmplă când sari de pe o stâncă. Datorită sentimentelor spontane care ne împing spre dulce și ne fac să evităm înălțimile, „pur și simplu știm" (la o primă aproximare) ce este de făcut și când. De exemplu, știm ce este de făcut când plâng bebelușii, când atacă prădătorii sau când ne ies în cale obstacole frustrante. Această cunoaștere înnăscută – care ne este transmisă explicit *numai* sub formă de sentimente – ne permite să supraviețuim în lumile foarte imprevizibile în care trăim, unde pe lângă noi gonesc autovehicule, iar aerul se umple de dioxid de carbon.

Așadar, pe măsură ce abandonăm iluzia familiară că conștiința circulă prin simțurile noastre, precum și concepția greșită conform căreia conștiința e sinonimă cu înțelegerea, să ne consolăm cu faptul că ea vine spontan din straturile noastre intime. Apare în noi chiar înainte să ne naștem. La sursa ei, suntem ghidați de un flux constant de sentimente ce provin dintr-un izvor de intuiție, apărut nu se știe de unde. Nu cunoaşte fiecare dintre noi, în mod individual, cauzele, dar le simțim. Sentimentele sunt o moștenire pe care ne-o lasă întreaga istorie a vieții, ca să ne călească în fața incertitudinilor viitoare.

Post-scriptum

La scurt timp după ce am scris o versiune completă a acestei cărți, am fost invitat să prezint teza principală în cadrul congresului anual Știința Conștiinței, care a avut loc la Interlaken, în 2019. Pentru asta a trebuit să condensez mai tot ce tocmai ai citit în formatul unei prelegeri plenare. Ar putea fi util să închei lunga noastră călătorie printr-un rezumat al celor 13 idei pe care le-am folosit pentru prelegerea respectivă.

(1) Marele fiziolog din secolul al XIX-lea Johannes Müller considera că organismele însuflețite „conțin un oarecare element nonfizic sau sunt guvernate de alte principii decât lucrurile neînsuflețite”. Studenții lui (Helmholtz, Brücke, Du Bois-Reymond, Ludwig și alții) nu erau de acord, ci erau siguri că „în organism nu acţionează nicio altă forță în afară de cele fizice și chimice obişnuite”. La rândul său, elevul lor Sigmund Freud a încercat să întemeieze o știință naturală a minții pe această bază, în care viața psihică poate fi redusă la „stări determinate cantitativ ale unor particule materiale detectabile”. A eșuat în acest proiect, lipsindu-i metodele, și l-a abandonat în 1896.

(2) Un secol mai târziu (1994), biologul inovator Francis

Crick a declarat că „tu, bucuriile și tristețile tale, amintirile și ambițiile tale, sentimentul identității tale personale și al liberului-arbitru, nu ești de fapt decât comportamentul unui mare ansamblu de celule nervoase, cu moleculele lor aferente”. Crick ne-a îndemnat să încercăm din nou să descoperim corelativele neuronale ale conștiinței, direcţie în care și-a îndreptat și el eforturile. Din păcate, a folosit însă conștiința vizuală ca model.

(3) În replică, filosoful David Chalmers a afirmat că preocuparea lui Crick pentru corelativele neuronale ale conștiinței era o problemă „ușoară” – una corelațională, și nu cauzală –, a cărei soluție putea să explice unde, dar nu și de ce și cum apare conștiința. Pentru Chalmers, problema „dificilă” a conștiinței era: cum și de ce activitățile neurofiziologice produc experiența conștiinței? Pentru el (și pentru predecesorul lui filosofic Thomas Nagel), problema se învârtea în jurul a ceva-cum-e al experienței: „Un organism are stări mentale conștiente dacă și numai dacă există ceva cum e să *fii* acel organism – ceva cum e *pentru* acel organism”. Prin urmare, problema dificilă este: de ce și cum apare calitatea subiectivă a experienței din evenimente neurofiziologice obiective?

(4) E cam vag să întrebi cum anume niște lucruri obiective *produc* niște lucruri subiective – asemenea termeni riscă să facă problema dificilă mai dificilă decât e necesar. Obiectivitatea și subiectivitatea sunt perspective observaționale, nu cauze și efecte. Evenimentele neurofiziologice nu pot să producă evenimente psihologice tot așa cum fulgerele nu pot să producă tunete. Ele sunt manifestări paralele ale unui singur proces de bază. Cauza fundamentală atât a fulgerului, cât și a tunetului

este electricitatea, ale cărei mecanisme legitime le explică pe amândouă. La fel, şi fenomenele fiziologice și psihologice pot fi reduse la cauze unitare, dar nu unele la celelalte.

(5) De obicei, descriem cauzele fundamentale ale fenomenelor biologice în termeni „funcționali”, iar mecanismele funcționale se pot reduce, la rândul lor, la legi naturale. De exemplu: care este mecanismul vederii? Totuși, Chalmers face observația corectă că mecanismul funcțional al vederii nu explică cum e să vezi. Asta pentru că vederea nu este o funcție intrinsec conștientă. Nu e nevoie ca îndeplinirea funcțiilor vizuale (nici chiar a celor specific umane, precum cititul) să se simtă în vreun fel. Percepția se întâmplă rapid, fără să conștientizezi ce este perceput, iar învățarea fără să conștientizezi ce se învață. Prin urmare, Chalmers a pus, pe bună dreptate, întrebarea: „De ce îndeplinirea acestor funcții este însoțită de experiență? De ce toată această prelucrare a informațiilor nu continuă «pe întuneric», fără să se simtă nimic în interior?”. Eșecul științei de a răspunde la această întrebare creează posibilitatea ca conștiința să nu facă parte din matricea cauzală obișnuită a universului.

(6) Întrebarea lui Chalmers poate fi pusă în legătură cu toate funcțiile cognitive, nu numai cu cele vizuale, dar nu și cu funcțiile afective. Cum poți să ai un sentiment fără să-l simți? Cum putem explica mecanismul funcțional al afectului fără să explicăm de ce și cum ne face să experiem ceva? Până și Freud era de acord în această privință: „Din esența unui sentiment face totuși parte faptul că el e resimțit, că e cunoscut, așadar, de către conștiință. Posibilitatea unui inconștient nu s-ar pune

absolut deloc atunci când vorbim de sentimente, senzații și afecte”.

(7) În acest context, e cât se poate de important să observăm că funcționarea corticală e însoțită de conștiință numai dacă este „facilitată” de sistemul reticulat activator din trunchiul cerebral superior. Leziuni de numai doi milimetri pătrați în această regiune distrug toată conștiința. Mulți cred că asta e din cauză că trunchiul cerebral modulează nivelul cantitativ al conștiinței sau „starea de veghe”, dar această opinie este neconvingătoare. Conștiința generată de trunchiul cerebral superior are un conținut calitativ propriu. Acesta este afectul. Întrucât conștiința corticală depinde de conștiința din trunchiul cerebral, rezultă că afectul este forma de bază a conștiinței. Subiectul capabil să simtă este constituit, concret, de afect.

(8) Afectul este o formă extinsă de homeostază, care este un mecanism biologic elementar ce a apărut în mod natural odată cu autoorganizarea. Sistemele cu autoorganizare supraviețuiesc deoarece ocupă stări limitate și nu se dispersează. Imperativul supraviețuirii a dus treptat la evoluția mecanismelor dinamice complexe aflate la baza intenționalității. Aspect esențial, ipseitatea sistemelor cu autoorganizare le conferă o perspectivă. Iată de ce are sens să discutăm despre subiectivitatea unui asemenea sistem: devierile de la stările lor viabile sunt consemnate de sistem, pentru sistem, ca nevoi.

(9) Afectul dă valențe hedonice nevoilor biologice, astfel încât deviațiile în creștere sau în scădere de la punctele de stabilizare homeostatice (erori de predicție în creștere sau în

scădere) sunt simțite ca neplăcere și, respectiv, plăcere. Fiecare categorie de nevoi – și sunt multe și diverse – are o calitate afectivă proprie și fiecare declanșează programe de acțiune preconizate să readucă organismul în limitele sale viabile. Aceste stări active – adică răspunsuri intenționale la stările afective – iau forma reflexelor și instinctelor înnăscute, care sunt completate gradual de învățarea din experiență, conform legii afectului[1]. Simțirea de către un organism a unor fluctuații ale propriilor nevoi permite realizarea unor alegeri și, astfel, susține supraviețuirea în contexte neprevăzute. Aceasta este funcția biologică a experienței.

(10) Nevoile nu pot fi simțite toate odată. Ele sunt prioritizate de un triunghi de decizie din mezencefal, unde nevoile curente (erorile de predicție reziduale, cuantificate ca energie liberă) ce converg în substanța cenușie periapeductală sunt clasificate în legătură cu oportunitățile curente (manifestate sub forma unei „hărți a predominanței” bidimensionale în coliculii superiori). Asta declanșează programe de acțiune condiționată, ce se desfășoară în contexte preconizate în cadrul unei ierarhii profunde a predicțiilor (modelul generativ al prozencefalului extins). Acțiunile generate de afecte prioritizate sunt voluntare, adică sunt supuse unor alegeri de tip aici-și-acum, și nu unor algoritmi prestabiliți. Asemenea alegeri sunt resimțite în conștiința exteroceptivă, care contextualizează afectul. Alegerile se fac pe baza unei cântăriri fluctuante a preciziei (numită și activare, modulare, amplificare postsinaptică) a semnalelor de eroare primite ce sunt făcute predominante de nevoile prioritizate, în timp ce sunt amortizate în memoria de lucru, cu scopul de a

minimiza incertitudinea (a maximiza încrederea) față de o predicție curentă referitoare la felul cum poate fi satisfăcută nevoia. Aceasta este „reconsolidarea”. După cum a spus Freud, „conștiința ia naștere acolo unde se sfârșește urma mnezică”.

(11) Alegerile de succes duc la ajustări pe termen lung ale predicțiilor senzorimotoare. Astfel, conștiința exteroceptivă este travaliu predictiv în desfășurare, al cărui scop este să stabilească predicții tot mai profunde (mai sigure, mai puțin conștiente) referitoare la modul în care se pot satisface nevoile. Această consolidare pe termen lung – și tranziția de la sisteme de memorie „declarativă” la cele de memorie „nondeclarativă” – necesită reducerea complexității modelului predictiv, pentru a facilita posibilitatea de generalizare. Oamenii aspiră la automatizare – încredere absolută –, dar nu o pot realiza niciodată în întregime. În măsura în care eșuăm, suportăm sentimente. Întrucât nu facem niciodată predicții perfecte, pulsiunea implicită (când totul merge bine) este CĂUTAREA – abordarea proactivă a incertitudinii, cu scopul de a o rezolva în avans. Când acest afect este prioritizat, îl resimțim drept curiozitate și interes față de lume.

(12) Acestea sunt mecanismele cauzale ale conștiinței – în toate manifestările ei, neurologice și psihologice deopotrivă –, aşa cum arată și cum se simte. Funcțiile fundamentale pot fi reduse la legi naturale, precum legea lui Friston[2]. Aceste legi stau la baza autoorganizării. Ele pot să explice cum și de ce resimțim opoziția proactivă față de entropie (adică uitare) în aceeași măsură în care alte legi științifice pot să explice alte lucruri naturale. Conștiința face parte din natură și poate fi

exprimată în termeni matematici.

(13) Toate sistemele conștiente cunoscute sunt vii, dar nu toate sistemele vii sunt conștiente. De asemenea, toate sistemele vii se autovalidează, dar nu toate sistemele care se autovalidează sunt vii. Dacă argumentul prezentat aici este corect, atunci, în principiu, putem construi cu ajutorul ingineriei un sistem care să se autovalideze și să fie conștient artificial. Conștiința poate fi *produsă*. Acest lucru va împlini cele mai fanteziste visuri ale lui Helmholtz și ale altor membri ai Societății de Fizică din Berlin. Totuși, trebuie să ne gândim ce ne motivează să-l facem, să acceptăm responsabilitatea colectivă pentru posibilele consecințele dezastruoase și să acţionăm cu cea mai mare prudență.

Anexă

Activare și informare

Într-o carte importantă pe tema activării cerebrale, Pfaff (2005, pp. 2-6) comentează:

> Satisfăcând nevoia unei „surse de energie” pentru comportament, activarea explică inițierea și persistența comportamentului motivat la o mare diversitate de specii [...]. Activarea, alimentând mecanismele pulsiunilor, potențează comportamentul, iar motivații și stimulente specifice explică de ce un animal face un anumit lucru, și nu altul [...]. *Dictionary of Ethology* nu numai că evidențiază activarea în contextul ciclului somn-veghe, ci se referă și la starea generală de receptivitate a animalului, după cum este indicată de intensitatea stimulării necesare pentru a declanșa o reacție comportamentală. Activarea „împinge animalul spre disponibilitatea de a acționa dintr-o stare de inactivitate”. În cazul acțiunii direcționate, unul dintre fondatorii etologiei, Niko Tinbergen, ar spune că activarea furnizează energia motrică pentru un „tipar de acțiune fixă” ca reacție la un „declanșator”. Dicționarul nu evită neurofiziologia, întrucât acoperă și nivelurile de activare indicate de electroencefalograma (EEG) corticală [...]. Generații de behavioriști au teoretizat și au confirmat experimental că un concept precum activarea este necesar pentru a explica inițierea, forța și persistența răspunsurilor comportamentale. Activarea furnizează forța fundamentală care face animalele și oamenii activi și receptivi astfel încât să manifeste comportamente instinctive sau învățate, dirijate către obiecte ale unor scopuri. Forța unui răspuns învățat depinde de activare și de pulsiune. Hebb a considerat starea de activare generalizată fundamentală pentru funcționarea cognitivă optimă. Duffy merge și mai departe, invocând conceptul „activării” pentru a

explica o parte semnificativă din comportamentul unui animal.

Analizele lui Pfaff asupra componentei principale sugerează că proporția comportamentului dintr-o serie amplă de date care poate fi explicat prin „activarea generalizată” este de 30-45%.

[Duffy] a anticipat că niște măsurători cantitative fiziologice sau fizice ar permite o abordare matematică a acestui aspect al științelor behavioriste [...]. Cannon a propus sistemul nervos autonom drept mecanism necesar prin care activarea pregătește animalul pentru acțiune musculară. Întregi teorii ale emoției s-au bazat pe activarea comportamentului [...]. Malmo a adunat toate aceste materiale, citând dovezi din EEG-uri și date fiziologice care sunt concordante cu rezultatele comportamentale în stabilirea activării și stimulării drept componentele principale ce impulsionează toate mecanismele comportamentale [...]. Aceasta este problema clasică a activării: care este modul în care influențele interne și externe trezesc creierul și comportamentul, la om și la alte animale, în laborator sau într-un cadru natural, etologic? E important să reformulăm și să rezolvăm această problemă deoarece e vorba de capacitatea de reacție la mediu, una dintre cerințele elementare pentru viața regnului animal. De asemenea, e oportun să o reformulăm și să o rezolvăm pentru că noi instrumente neurobiologice, genetice și computaționale oferă abordări ale „stărilor comportamentale” ce nu au mai fost posibile până acum [...]. Explicarea activării ne va permite să înțelegem stările comportamentale ce se află la baza unui număr mare de mecanisme specifice de reacție. Nu numai că e strategic să analizezi mai multe comportamente odată, ci și elucidarea mecanismelor unor stări comportamentale duce la o înțelegere a dispoziției și temperamentului. Altfel spus, o bună parte din neuroștiințele secolului XX au vizat explicarea particularității unor conexiuni specifice stimul-răspuns. Acum avem cum să dezvăluim mecanismele unor clase întregi de răspunsuri sub denumirea de „control al stării”. Cele mai importante sunt mecanismele ce determină nivelul de activare [...]. Orice definiție cu adevărat universală a activării trebuie să fie elementară și fundamentală, primitivă și nediferențiată și să nu derive din funcții superioare ale SNC. Nu poate fi limitată de condiții sau criterii temporare, specifice. De exemplu, nu se poate limita la explicarea răspunsurilor la un singur tip de senzație. Trebuie incluse și activitatea motoare voluntară și răspunsurile emoționale. Așadar, propun următoarea definiție operațională care este satisfăcătoare la un nivel intuitiv și care va duce la măsurători cantitative precise: *„activarea generalizată” este mai mare la un animal sau la un om care este: (S) mai atent la stimuli senzoriali de toate*

tipurile, (M) mai activ sub aspect motric și (E) mai reactiv din punct de vedere emoțional. Aceasta este o definiție concretă a forței fundamentale din sistemul nervos [...]. Toate cele trei componente pot fi măsurate cu precizie [...]. În mod clar, există o neuroanatomie a activării generalizate, există neuroni ale căror tipare de descărcare duc la ea, precum și gene a căror pierdere o perturbă. Prin urmare [...], activarea generalizată este starea comportamentală produsă de căile de activare, de mecanismele electrofiziologice și de influențele genetice. Faptul că aceste mecanisme produc aceleași atenție senzorială (S), reactivitate motoare (M) și reactivitate emoțională (E) ca în definiția noastră demonstrează existenţa unei funcții de activare generalizată, precum și acuratețea definiției sale operaționale.

Pfaff continuă: „Deoarece activarea SNC depinde de surprindere și de impredictibilitate, cuantificarea ei corectă depinde de matematica *informațiilor*" (p. 13, sublinierea mea). Ecuația lui Shannon (1948) face informațiile măsurabile, după cum explică Pfaff:

Dacă orice eveniment este perfect regulat – de pildă, bătaia unui metronom –, evenimentul următor (bătaia următoare) nu ne spune nimic nou. Are o probabilitate extrem de ridicată (*p*) de a apărea în exact acel interval de timp (*time bin*) [...]. Nu avem nicio incertitudine că va avea loc bătaia, indiferent de *time bin*. În ecuația lui Shannon, informațiile din orice eveniment sunt invers proporționale cu probabilitatea lor. Altfel spus, cu cât suntem mai nesiguri în privința apariției evenimentului respectiv, cu atât mai multe informații se transmit, în mod inerent, când acesta se întâmplă [...]. Când toate evenimentele dintr-o serie de evenimente sunt la fel de probabile, informațiile au valoarea maximă. Dezordinea maximizează fluxul de informații. Termenul tehnic, împrumutat din termodinamică, pentru dezordine în ecuația lui Shannon este *entropie*. Simbolul lui Shannon pentru entropie este *H* [...]. Conținutul de informație inerent într-un eveniment *x* oarecare este:

$$H(x) = p(x)\log_2\frac{1}{p(x)}$$

unde $p(x)$ este probabilitatea evenimentului x.
Pfaff rezumă (pp. 19-20):

Pentru ca un animal inferior sau un om să fie activat, trebuie să existe schimbare

în mediul său [interoceptiv sau exteroceptiv]. Dacă există schimbare, trebuie să existe incertitudine cu privire la starea mediului. Din punct de vedere cantitativ, în măsura în care există incertitudine, predictibilitatea scade. Având în vedere aceste considerații, putem folosi [ecuația lui Shannon] ca să afirmăm că, pe cât mediul este mai puțin predictibil și entropia este mai mare, pe atât există mai multe informații. Activarea creierului și comportamentului, precum și calculele informaționale sunt unite inseparabil.

Pe scurt, stimulii necunoscuți, nepreconizați, dezordonați și neobișnuiți (cu multe informații) produc și susțin răspunsuri de activare (p. 23).

Teoria informației s-a ascuns în tot acest timp în spatele unor cercetări comportamentale și date neurofiziologice. Mai întâi, conform unei logici simple și clare, gândește-te la ce este necesar pentru ca un animal sau un om să se hotărască să acționeze. În al doilea rând, gândește-te la ce este necesar ca să recunoști un stimul familiar (obișnuința) și ca să acorzi o atenție specială unui stimul nou. În al treilea rând, din punctul de vedere al experimentatorului, teoria informației furnizează metode pentru calcularea conținutului semnificativ al așa-numitelor *spike trains* și pentru cuantificarea încărcăturii cognitive a anumitor situații de mediu. Se pot pune întrebări noi: de câtă distorsiune a unui câmp de stimuli senzoriali e nevoie pentru noutate? Ce tipuri de generalizare din partea unui anumit tip de stimul sunt permise pentru un anumit tip de răspuns? Abordarea teoriei informației ne va ajuta să transformăm combinaţia dintre genetică, neurofiziologie și comportament într-o știință cantitativă. Putem folosi „matematica activării” ca să ne ajute în analiza mecanismelor neurobiologice.

În cele din urmă, Pfaff concluzionează (pp. 138-145):

Sistemele de activare ale SNC se luptă eroic cu legea a doua a termodinamicii într-un mod foarte special. Ele răspund selectiv la situații de mediu ce au o entropie inerent mare – un grad ridicat de incertitudine și, prin urmare, de conținut informațional. Dar, când răspund, sistemele de activare ale SNC reduc eficient entropia prin comprimarea tuturor acelor informații într-un singur răspuns legitim [...]. Neurobiologia activării este neuroștiința schimbării, incertitudinii, impredictibilității și surprizei – adică a științei informației. Pe parcursul tuturor analizelor de până acum asupra mecanismelor activării din SNC – neuroanatomice, fiziologice, genetice și comportamentale –, conceptele teoriei

informației se dovedesc a fi utile. Matematica informației furnizează căi pentru clasificarea răspunsurilor la stimulii naturali. De fapt, celulele nervoase codifică probabilități și incertitudini, rezultând că ele pot să ghideze comportamentul în circumstanțe nepredictibile. Însăși activarea SNC depinde cu totul de schimbare, incertitudine, impredictibilitate și surpriză. Fenomenul numit obișnuință, o scădere a amplitudinii răspunsului la repetarea aceluiași stimul, se întâlnește peste tot în neurofiziologie, științele comportamentale și fiziologia SNA, arătându-ne în ce mod conținutul informațional în scădere duce la scăderea activării SNC. Astfel, teoria activării și teoria informației au fost făcute una pentru cealaltă.

E important să recunoaștem că „matematica informației" explică comportamentul neuronilor atât în procesele de activare, cât și în cele de învățare, ce, combinate, determină ceea ce *face* creierul. Așadar, deși „informația" nu este un construct fiziologic, ea explică în mod legitim activitatea fiziologică a creierului. Ea e *funcția* selectată de evoluție, iar fenotipurile fiziologice vin după.

Mulțumiri

Le sunt recunoscător următorilor prieteni și colegi pentru că au citit mai multe versiuni ale capitolelor: Richard Astor, Nikolai Axmacher, Samantha Brooks, Aimee Dollman, George Ellis, Karl Friston, Eliza Kentridge (care este mult mai mult decât o prietenă), Joe Krikler, Joshua Martin, Lois Oppenheim, Jonathan Shock, Pippa Skotnes și Dawie van den Heever. Îi sunt deosebit de îndatorat lui Ed Lake pentru că a făcut manuscrisul mult mai accesibil lecturii – nu am mai văzut redactor care să lucreze atât de temeinic. Trevor Horwood a făcut corectura manuscrisului, cu o contribuție din partea lui Tim James.

Aș vrea să-i mulțumesc și lui sir Sydney Kentridge pentru că mi-a permis să-i folosesc casa din Chailey, unde am scris cea mai mare parte a cărții în iernile din 2018-2019 și 2019-2020. Ca de fiecare dată, în culise au fost bravele mele asistente, Paula Barkay și Eleni Pantelis. Această carte, ca mai tot ce am făcut până acum, nu s-ar fi întâmplat fără ele.

Note

Capitolul 1. Materia viselor

1. Popper (1963). E adevărat că nu toată lumea este de acord cu această formulare despre cum funcționează știința. Cu toate acestea, aproape toți specialiștii în științe naturale subscriu la ea. *Acum, că citești aceste note, se impune o explicație. Notele se adresează în principal cititorilor din mediul academic care sunt interesați de (sau sunt familiarizați cu) diversele lucrări de specialitate cu care se înrudește această carte. Notele pot fi ignorate fără probleme de publicul larg – principalii cititori cărora mă adresez*.
2. Freud (1893a), p. 13.
3. Sacks (1984), p. 164.
4. Freud (1895), p. 160. După vreo 36 de ani, Freud i-a scris într-un stil emoționant lui Albert Einstein despre lipsa prestigiului științific al psihologiei prin comparație cu fizica (Freud, 1994, p. 239): „Într-adevăr, nu e cu totul regretabil să optezi pentru psihologie. Nu există o temă mai măreață, mai bogată și mai misterioasă, demnă de fiecare efort al intelectului uman, decât viața minții. Cu siguranță, psihologia este cea mai frumoasă dintre toate doamnele

nobile, doar că prințul ei e condamnat să rămână nefericit în dragoste”.

5. „*Ador* acest pasaj din Freud – sunt foarte fericit că l-ai găsit. După cum spui, cu generozitate, ceva analog se poate spune şi despre propriile mele istorice de caz și despre istoricele unor cazuri neurologice (cel puțin neuropsihologice) în general. L-am citat (deși nu știu dacă va supraviețui – manuscrisul meu a devenit mult prea lung și cu prea multe note de subsol) într-un text tocmai finalizat, mai degrabă general – «Scotoma» –, despre uitare și neglijare în știință.” (Scrisoare de la Sacks, datată 2 ianuarie 1995)
6. După cum a scris la vremea respectivă Semir Zecki, specialist în neuroștiințe: „Majoritatea dintre noi s-ar îngrozi la gândul de a investiga ceea ce pare să fie o problemă atât de impenetrabilă” (Zeki, 1993, p. 343).
7. Aserinsky și Kleitman (1953).
8. Dement și Kleitman (1957).
9. Vezi Freud (1912), p. 259: „Există un produs psihic care se poate întâlni la persoanele cele mai normale și care oferă totuși o analogie extrem de izbitoare cu manifestările cele mai sălbatice ale nebuniei, produs ce nu a fost mai inteligibil pentru filosofi ca nebunia însăși. Mă refer la vise. Psihanaliza se bazează pe analiza visului; interpretarea visului este partea cea mai completă a muncii pe care a realizat-o până astăzi tânăra știință” [*Opere esențiale*, vol. III: *Psihologia inconștientului,* „Câteva observaţii asupra conceptului de inconştient în psihanaliză”, traducere din limba germană de G. Lepădatu, Trei, București, e-book].
10. Popper (1963).
11. Din punct de vedere strict anatomic, talamusul nu este

considerat parte a trunchiului cerebral. Totuși, din punct de vedere fiziologic, unii dintre nucleii săi „nonspecifici” fac parte din sistemul reticulat activator; prin urmare, sunt grupați alături de funcțiile trunchiului cerebral (de unde și sintagma „sistem activator reticular-talamic extins”, ERTAS). Nucleii talamici „specifici”, care acționează în principal ca stații de transmitere a semnalelor senzoriale, sunt grupați împreună cu funcțiile cortexului. În această carte voi folosi termenii „trunchi cerebral” și „cortex” în principal pentru a desemna deosebirea funcțional-anatomică dintre *stimularea* ERTAS și, respectiv, *reprezentarea* talamo-corticală. De aceea, voi considera structuri ale trunchiului cerebral nu numai talamusul și hipotalamusul „nonspecifici”, ci și nucleii prozencefalului bazal. Pentru o viziune contemporană asupra sistemului activator reticular-talamic *extins*, vezi Edlow *et al.* (2021).

12. Jouvet (1965).
13. Hobson, McCarley și Wyzinski (1975).
14. McCarley și Hobson (1977), p. 1346.
15. *Ibidem*, p. 1219.
16. Singurul pacient care a spus că nu mai visa a suferit aproape sigur leziuni semnificative mult dincolo de tegmentul mezopontin, care generează somnul REM (din cauza unei hemoragii subarahnoidiene traumatice; Lavie *et al.*, 1984). Prin urmare, modificările la nivelul capacității sale de a visa erau dificil de corelat cu o anumită arie cerebrală.
17. Vezi Solms (1997a) pentru o descriere completă a concluziilor mele. Mi-am depus teza în 1991, dar am publicat-o abia după șase ani.

18. Solms (2000a). Principiul „dublei disocieri” în neuropsihologie ne permite să diferențiem funcțiile mintale la punctele lor de întâlnire: dacă o leziune în aria *X* (a creierului) cauzează pierderea funcției *A*, dar nu și a funcției *B*, iar o leziune în aria *Y* cauzează pierderea funcției *B*, dar nu și a funcției *A*, atunci funcția *A* și funcția *B* nu pot fi același lucru. Cu alte cuvinte, în acest caz, funcția somnului REM și funcția visului nu pot fi identice. Ele *se corelează* reciproc (adică apar în același timp), dar nu sunt același lucru.
19. Multe alte dovezi, pe lângă constatările mele despre leziuni, susțin această concluzie. De exemplu, există o șansă de 50% de a obține relatări ale unor vise în primele câteva minute ale somnului (în faza 2 descendentă), cu mult înainte de primul episod REM. De asemenea, visele care nu se pot deosebi deloc de cele REM apar în somnul non-REM cu o frecvență în creștere în faza matinală de trezire a ritmului diurn, ceea ce se numește „efectul dimineții târzii”. De asemenea, deși visul este mult mai frecvent în somnul REM decât în somnul non-REM, somnul non-REM este dominant în comparație cu somnul REM. În consecință, cel puțin un sfert dintre toate visele apar în timpul somnului non-REM. Vezi Solms (2000a) pentru detalii.
20. Solms (1991, 1995).
21. Frank (1946, 1950), Partridge (1950).
22. Schindler (1953).
23. Hartmann *et al.* (1980).
24. Sharf *et al.* (1978). Studiile ulterioare au arătat că antagoniștii dopaminei au efectul opus (Yu, 2007).
25. Dahan *et al.* (2007).

26. Léna *et al.* (2005).
27. Solms (2011).
28. Solms (2001).
29. Acesta denumire nu e bună, e o revenire la vremurile behavioriste. Există multe tipuri diferite de „recompensă” (adică plăcere) în creier.
30. Rolls (2014), Berridge (2003), Panksepp (1998).
31. Panksepp (1998), p. 155.
32. Activitatea de CĂUTARE dopaminergică (spre deosebire de alte activități ale monoaminelor) continuă după instalarea somnului și este la nivel maxim în timpul somnului REM. Probabil nu e o coincidență că coincide cu mișcările sacadate rapide ale ochilor. Mișcările ochilor la oameni, la fel ca adulmecatul și mișcările mustăților la rozătoare, sunt un bun echivalent al activării CĂUTĂRII (vezi Panksepp, 1998).
33. Pace-Schott și Hobson (1998).
34. Braun (1999), p. 196.
35. *Ibidem*, p. 201.

Capitolul 2. Înainte de Freud și după

1. Atunci am fost acceptat la specializare, dar nu am început-o până în 1989.
2. De fapt, Freud nu a pus un titlu acestui manuscris nepublicat; titlul a fost inventat de traducători. În corespondența cu Wilhelm Fliess, Freud l-a numit „Psihologie pentru neurologi”, „Schiță de psihologie” și „Psihologia”.
3. Scrisoare către Hallman (1842), publicată în du Bois-

Reymond (1918), p. 108. De asemenea, citată frecvent este prefața lui du Bois-Reymond la lucrarea sa *Über die Lebenskraft* (1848-1884, pp. xliii-xliv): „Nu există forțe noi ce funcționează în organisme și în particulele lor, alte forțe care să nu funcționeze și în afara lor. De asemenea, nu există nici forțe care merită numele «forță vitală». Separarea dintre așa-numitele naturi organică și anorganică este complet arbitrară”.

4. Bechtel și Richardson (1998).
5. Vezi comentariile mele despre acest termen în Solms (în curs de apariție).
6. Freud (1950b), p. 295.
7. Prioritatea lui Freud în formularea poziției „funcționaliste” nu este larg recunoscută (Freud, 1900, p. 536): „[Încercăm să înțelegem] gradul de complicație a activității psihice, prin descompunerea ei, și indicarea funcționării individuale a părților individuale ale aparatului. După câte știu, nu s-a încercat încă ghicirea alcătuirii instrumentului psihic dintr-o astfel de descompunere. Mie mi se pare inofensivă” [Sigmund Freud, *Opere esențiale*, vol. 2: *Interpretarea viselor*, traducere din limba germană de Roxana Melnicu. București: Trei, 2010, pp. 614-615]. *Cf* Shallice (1988).
8. Când Freud a introdus pentru prima dată acest termen ciudat, a explicat că se referă la un nivel al explicației ce încorporează atât psihologia, cât și biologia (scrisoare către Fliess din 10 martie 1989; Freud, 1950a). După cum i-a scris o dată Freud lui Georg Groddeck: „Inconștientul este îndelung căutata verigă lipsă dintre fizic și mintal” (scrisoare din 5 iunie 1917). Vezi Solms (2000b). Vezi și prezentarea mea de la o reuniune a Academiei de Științe din New York

organizată pentru a sărbători centenarul „Proiectului” lui Freud (Solms, 1998). Marele Karl Pribram a luat cuvântul la reuniunea respectivă. Tot acolo l-am întâlnit și pe Joseph Bogen, un neurofiziolog inovator. Mi-l amintesc clar afirmând cu lejeritate că conștiința este generată de nucleii intralaminari ai talamusului, aceasta fiind prima dată când am auzit pe cineva sugerând că cortexul nu este conștient în mod intrinsec. Vezi Bogen (1995).

9. Critica adusă de Freud (1891) localizaționismului a pus bazele abordării „sistemelor funcționale” care a dominat mai târziu neuropsihologia și apoi cognitivismul. Am discutat în detaliu despre aceste lucruri în Solms și Saling (1986) și Solms (2000b).
10. De unde și titlul cărții lui Sulloway (1979), *Freud: Biologist of the Mind.*
11. Freud (1914), p. 78.
12. Freud (1920), p. 83.
13. Scrisoare către Fliess, 20 octombrie 1895.
14. Scrisoare către Fliess, 29 noiembrie 1895.
15. Scrisoare către Fliess, 25 mai 1895.
16. În prezent pregătesc o traducere în engleză a operelor complete ale lui Sigmund Freud legate de neuroștiințe (în patru volume). Vezi și ediția revăzută de mine a traducerilor lui Strachey, *The Revised Standard Edition of the Complete Psychological Works of Sigmund Freud* (24 de volume).
17. Freud (1950b), pp. 303, 316. Cu mulți ani înaintea lui Freud, Baruch Spinoza a scris că „dorința este natura însăși sau esența unei persoane”.
18. Freud (1901), p. 259.
19. Freud (1920), p. 60.

20. Freud (1915a), pp. 121-122, sublinierea mea.
21. Scrisoare către Fliess, 25 mai 1895.
22. Freud (1940), p. 197.
23. Această figură este adaptată după Braun *et al.* (1997). Studiul lui Braun a fost pur descriptiv. Constatările lui sunt compatibile cu teoria lui Freud, dar nu o confirmă experimental deoarece nu au testat nicio predicție derivată din ea. Totuși, o studentă a mea (Catherine Cameron-Dow, 2012) a testat recent în mod direct teoria lui Freud că visele protejează somnul. Confirmarea de către ea a ipotezei este acum urmărită într-un studiu mai amplu al colegei mele Tamara Fischmann, în Berlin.
24. Întâmplător, dezbaterea a fost prezidată de nimeni altul decât David Chalmers. Rezultatul a inversat un vot din 1978 ce urmase prezentării de către Hobson a teoriei sale despre „activare-sinteză” în plenul Asociației Americane de Psihiatrie.
25. Solms și Saling (1986).
26. Vezi Braun (1999).
27. Malcolm-Smith *et al.* (2012).
28. Solms (2000c), Solms și Zellner (2012).
29. E interesant că Freud și-a elaborat conceptul de „pulsiune libidinală” într-o perioadă în care consuma regulat cocaină, un alcaloid ce activează intens sistemul CĂUTĂRII dopaminergic. Oare e prea fantezist să presupunem că experiența personală a lui Freud cu efectele motivaționale generalizate ale cocainei a contribuit la faptul că a recunoscut existența în minte a unui asemenea mecanism motivațional universal?
30. În capitolul 7 voi arăta legătura dintre „împlinirea

dorințelor” și codificarea predictivă și cea dintre „testarea realității” și ceea ce numim astăzi „eroare de predicție” (sau eroare de predicție modulată de precizie).

31. Fotopoulou, Solms și Turnbull (2004).
32. Turnbull, Jenkins și Rowley (2004).
33. Fotopoulou și Conway (2004), Turnbull, Berry și Evans (2004), Fotopoulou *et al.* (2007, 2008a, 2008b), Turnbull și Solms (2007), Fotopoulou, Conway și Solms (2007), Fotopoulou (2008, 2009, 2010a, 2010b), Coltheart și Turner (2009), Cole *et al.* (2014), Besharati, Fotopoulou și Kopelman (2014), Kopelman, Bajo și Fotopoulou (2015).
34. Vezi Turnbull, Fotopoulou și Solms (2014) pentru recenzii. Vezi și Besharati *et al.* (2014, 2016).
35. Zellner *et al.* (2011).
36. Solms și Turnbull (2002, 2011), Panksepp și Solms (2012), Solms (2015a).
37. În cele din urmă, aceste prezentări au fost adunate într-un volum: Kaplan-Solms și Solms (2000).
38. Kandel (1998).
39. Kandel (1999), p. 505.
40. Printre specialiștii în neuroștiințe se numărau Allen Braun, Jason Brown, Antonio Damasio, Vittorio Gallese, Nicholas Humphrey, Eric Kandel, Marcel Kinsbourne, Joseph LeDoux, Rodolfo Llinás, Georg Northoff, Jaak Panksepp, Michael Posner, Vilanayur Ramachandran, Oliver Sacks, Todd Sacktor, Daniel Schacter, Carlo Semenza, Tim Shallice, Wolf Singer și Max Velmans. Printre psihanaliști – Peter Fonagy, Andre Green, Ilse Grubrich-Simitis, Otto Kernberg, Marianne Leuzinger-Bohleber, Arnold Modell, Barry Opatow, Allan Schore, Theodore Shapiro, Riccardo Steiner și Daniel

Widlöcher.

41. Damasio, Damasio și Tranel (2013).

42. Freud (1940), p. 198. Pe tot parcursul acestei cărți, folosesc versiunile revăzute ale traducerilor lui James Strachey (vezi Solms, în curs de apariție).

43. Freud (1920), p. 24, sublinierea mea.

44. Freud (1940), pp. 161-162, sublinierea mea.

45. Vezi Freud (1923), p. 26: „Eul este înainte de toate o entitate corporală, nu doar una situată la suprafață, ci însăși proiecția unei suprafețe. Pentru a ne servi de o analogie anatomică, el poate fi comparat cu «manechinul cerebral» al anatomiștilor, situat în scoarța cerebrală: cu capul în jos și picioarele în sus, ochii întorși spre înăuntru și având zona limbajului în stânga [...]. În acest fapt avem o nouă demonstrație a ceea ce afirmam mai sus în legătură cu Eul conștient, și anume că el reprezintă în primul rând un Eu corporal" [Sigmund Freud, „Eul și Se-ul", traducere de G. Purdea și Vasile Dem. Zamfirescu, în *Opere esenţiale*, vol. 3: *Psihologia inconștientului*. București: Trei, e-book].

46. Vezi Solms (2013).

Capitolul 3. Eroarea privitoare la cortex

1. Merker (2007), p. 79. Aceste observații au confirmat un raport anterior întocmit de Shewmon, Holmes și Byrne (1999).

2. Relatarea ce urmează despre constatările lui Merker este parafrazată după raportul pe care l-a publicat (2007).

3. Damasio și Carvalho (2013), p. 147.

4. Aceste observații parafrazează rezumatul lui Merker asupra

literaturii de specialitate (2007), p. 74.

5. Panksepp (1998).

6. Vezi Weiskrantz (2009).

7. Urmărește videoclipul: https://blogs.scientificamerican.com/observations/blindsight-seeing-without-knowing-it/

8. Același lucru e valabil şi în cazul căii de la ochi la corpul geniculat lateral. Vezi figura 6.

9. Zeman (2001).

10. Coenen (2007), p. 88, sublinierea mea.

11. Nu există *răspuns ocular* = 1 punct, deschide la stimul de durere = 2 puncte, deschide la ordin verbal = 3, deschide spontan = 4; nu există *răspuns verbal* = 1, răspuns cu sunete neinteligibile = 2, răspuns cu cuvinte fără legătură = 3, răspuns confuz = 4, răspuns orientat = 5; nu există *răspuns motor* = 1, extensie (decerebrare) = 2, flexie anormală (decorticare) = 3, flexie normală (răspuns de retracție) = 4, localizarea durerii = 5, răspunde la comenzi verbale = 6.

12. Se afirmă și că iluzia ia naștere din „eroarea moralistă”. Heather Berlin, specialistă în neuroștiințe cognitive, a calificat drept „arbitrar” să presupui că abilitatea de a răspunde la stimuli externi implică conștiință internă (Berlin, 2013, pp. 25-26): „Ipoteza principală a lui Solms conform căreia copiii cu hidroanencefalie sunt conștienți este neîntemeiată. Nu putem presupune că a avea un ciclu de somn-veghe și expresii ale emoțiilor (râs, furie etc.) reclamă conștiință [...]. Deși este adevărat că ei pot fi de fapt conștienți, nu putem presupune că și sunt. Procesele inconștiente pot fi destul de sofisticate și de complexe (Berlin, 2011). Esența teoriei lui Solms se întemeiază pe o proiecție a existenței unei conștiințe bazate pe ceea ce par a

fi comportamente emoționale semnificative, un exemplu de «eroare moralistă» (să afirmi că ceva trebuie să fie adevărat pentru că ne face să ne simțim bine să credem afirmația respectivă). Oamenii doresc în mod natural să presupună că există conștiință”.

Iată ce am scris eu în replică (Solms, 2013, pp. 80-81): „De ce ar trebui să presupunem că manifestările emoționale adecvate contextual, care sunt imediat evocate de stimularea unei anumite regiuni cerebrale și anulate de leziuni ale aceleiași regiuni cerebrale și care corespund trăirilor noastre afective, *nu* corespund unor trăiri afective ale acestor copii și animale? Cu siguranță că această presupunere ar fi mai «arbitrară» decât a mea. Singura dovadă în acest sens este că acești copii și animale nu își pot «declara» sentimentele *în cuvinte*”.

13. Cunoscute sub denumirea de legile de asociație: prin contiguitate, repetiție, atenție, similaritate etc.

14. Apercepția este „procesul prin care experiența nouă este asimilată și transformată de restul experiențelor trecute ale unei persoane pentru a forma un nou întreg” (Runes, 1972).

15. Meynert (1867).

16. Această afirmație omite în mod convenabil faptul că creierul este mai mare la alte mamifere (precum elefanții). Cortexul uman nu este mai mare decât cel al altor mamifere nici măcar în privința proporției dintre cortex și mărimea corpului sau dintre cortex și formațiunile subcorticale.

17. Campbell (1904), pp. 651-652.

18. Meynert (1884; traducere în engleză, 1885), p. 160.

19. Munk (1878, 1881).

20. S-a considerat că baza anatomică pentru deosebirea dintre

orbire și cecitatea minții este faptul, descoperit de Paul Flechsig (1901, 1905), că cortexul striat „de proiecție” conține celule primordiale conectate direct la periferia retiniană. Aceste celule primordiale sunt mielinizate la naștere și, prin urmare, nu conțin imagini *mnezice*. Cortexul „de asociere” din jur – vehiculul tuturor funcțiilor psihice – se mielinizează mult mai târziu. Afirmația e valabilă în cazul celorlalte regiuni corticale specifice inputurilor senzoriale.

21. Wilbrand (1887, 1892). Vezi traducerea făcută de mine în engleză a studiului de caz original al lui Wilbrand, unde multe dintre aceste idei teoretice sunt discutate detaliat (Solms, Kaplan-Solms și Brown, 1996).

22. Această observație a lui Wilbrand pleca de la o observație similară anterioară a lui Charcot (1883) despre un pacient care avea vise *nonvizuale*, astfel apărând conceptul de „sindrom Charcot-Wilbrand”: incapacitatea de a revizualiza și a recunoaște obiecte vizuale ziua și pierderea capacității de a visa noaptea. Vezi Solms, Kaplan-Solms și Brown (1996) și Solms (1997a) pentru o analiză critică a conceptului Charcot-Wilbrand în contextul concluziilor mele ulterioare.

23. În consecință, a făcut o distincție între surditatea comună și „surditatea verbală” (afazie). Despre afazia lui Wernicke se spunea că este cauzată de leziuni la nivelul regiunii corticale ce conține imagini mnezice auditive pentru cuvinte – amintirea sunetelor vorbirii –, iar surditatea comună apărea în urma unor leziuni la nivelul căilor subcorticale ce conectează această regiune de senzațiile auditive din exterior. Paul Pierre Broca (1861, 1865) descrisese anterior o formă paralelă de afazie, cauzată de afectarea imaginilor motoare pentru cuvinte – adică afectarea programelor

deprinse despre modul în care *produci* sunetele vorbirii. Conform unui student al lui Wernicke, Ludwig Lichtheim (1885), alte forme de afazie erau cauzate de afectarea căilor transcorticale „de asociere” ce duc de la imaginile mnezice auditive și motoare pentru cuvinte la idei *abstracte*, ceea ce conferea imaginilor concrete semnificația lor. Heinrich Lissauer (1890) a subdivizat în mod similar „cecitatea psihică” în două tipuri, un tip „aperceptiv”, cauzat de afectarea imaginilor mnezice vizuale, și un tip „asociativ”, cauzat de afectarea căilor transcorticale ce duc de la imaginile mnezice pentru obiecte vizuale la cele pentru idei abstracte. Freud (1891) a fost cel care, mai târziu, a redenumit cecitatea psihică drept „agnozie” vizuală. Apoi, Hugo Liepmann (1900) a împărțit în mod similar „paralizia psihică” (apraxia) în tipuri aperceptive și asociative – apraxie „chinestezică a membrelor” și, respectiv, „ideomotoare” –, după cum s-a menționat în text.

24. Liepmann (1990).

25. În prefață a scris: „Cititorul nu va găsi o altă definiție a «psihiatriei» în această carte în afară de cea dată pe pagina de titlu: Tratat clinic de boli ale prozencefalului. Denumirea istorică pentru psihiatrie, adică «tratarea sufletului», implică mai mult decât putem să realizăm și transcende granițele investigării științifice riguroase”.

26. Vezi Absher și Benson (1993) și Goodglass (1986). E interesant că Antonio Damasio i-a fost student lui Geschwind.

27. Se întâmplă că știm că pacienții cu afecțiuni corticale și fără limbaj au conștiența nealterată, deoarece pot să-și comunice sentimentele în alte feluri și o și fac. Vezi Kaplan-Solms și

Solms (2000) pentru descrieri detaliate ale unor rapoarte introspective nonverbale făcute de pacienți afazici de diferite tipuri. După o oarecare ezitare inițială în secolul al XIX-lea, consensul general a fost întotdeauna că pierderea limbajului nu afectează „inteligența” în mod semnificativ.

28. Merker (2007), p. 65.
29. *Cf.* distincția făcută de Ned Block (1995) între conștiința „fenomenologică” și „de acces”.
30. Vezi Craig (2009, 2011).
31. Vezi Dehaene și Changeux (2005); Baars (1989, 1997). Sintagma „gânduri de ordin superior” este din Rosenthal (2005).
32. Qin *et al.* (2010), Mulert *et al.* (2005).
33. Dialogul este extras dintr-un raport verbal al cazului prezentat la Congresul de Neuropsihanaliză din Berlin în 2011. Studiul de caz a fost publicat ulterior în Damasio, Damasio și Tranel (2013).
34. Damasio, Damasio și Tranel (2013).
35. Harlow (1868).
36. LeDoux și Brown (2017).
37. Vezi LeDoux (1999), p. 46, sublinierea mea: „Atunci când stimulii electrici aplicați pe amigdala umană generează sentimente de frică (vezi Gloor, 1992) nu e din cauză că amigdala «simte» frica, ci din cauză că diversele rețele activate de amigdală furnizează memoriei funcționale inputuri care sunt *etichetate* drept frică”.
38. Whitty și Lewin (1957), p. 73.
39. Solms (1997a), p. 186, Cazul 22.
40. Epitetul a fost introdus de Karl Jaspers (1963).

Capitolul 4. Care este obiectul experienței?

1. Vezi Solms și Saling (1990) pentru o discuție detaliată despre dezacordurile dintre Freud și Meynert și alți specialiști în neuropsihologie de la vremea respectivă. Inițial, și-a prezentat opiniile separate într-un manuscris nepublicat (Freud, 1887). Acestea au fost tipărite pentru prima dată în Freud (1888). Apoi le-a reiterat în cartea sa din 1891 și le-a dezvoltat în Freud (1893b) și, în cele din urmă, în scrisoarea către Fliess din 6 decembrie 1896.
2. Freud (1891), traducerea mea. Freud (1886), p. 14, a raportat că a făcut „vizite repetate” la laboratorul lui Munk din Berlin. Traseul căilor neuronale din trunchiul cerebral constituia o preocupare aparte a lui Freud în studiile sale de anatomie, în anii 1880.
3. Freud (1891), traducerea mea.
4. *Ibidem*, traducerea mea. Freud a pus sub semnul întrebării și baza anatomică și fiziologică a distincției făcute de Meynert și de alții între etapele aperceptive și cele asociative ale procesării corticale (*ibidem*).
5. Scrisoare către Wilhelm Fliess, datată 6 decembrie 1896 (Freud, 1950a, p. 233).
6. Prin urmare, la Freud, sistemele inconștiente și preconștiente coincid cu ceea ce numim astăzi memorie pe termen lung „nondeclarativă” și „declarativă”, iar sistemul conștient coincide cu ceea ce numim memorie pe termen scurt. Prima și ultima dintre cele cinci etape ale lui Freud (percepția și conștiința) sunt *stări* ale neuronilor, și nu urme.
7. Kihlstrom (1996).
8. Bargh și Chartrand (1999), p. 476.

9. Squire (2009).
10. Claparède (1911).
11. Galin (1974), p. 573.
12. Vezi McKeever (1986).
13. Crick și Koch (1990), Newman și Baars (1993), Dehaene și Naccache (2001), Bogen (1995), Edelman (1990), Marc și Llinás (1994), Tononi (2012).
14. Nagel (1974). Vezi și Chalmers (1995a) și Strawson (2006).
15. Freud (1915b), p. 177, explică: „Toată deosebirea [dintre idei și afecte] provine din faptul că reprezentările sunt investiri – în mod fundamental, ale urmelor mnezice –, în timp ce afectele și sentimentele corespund proceselor de descărcare, ale căror ultime manifestări sunt percepute ca senzații. În situația actuală a cunoștințelor noastre despre afecte și sentimente, nu putem enunța mai clar această deosebire” [Sigmund Freud, *Opere esențiale*, vol. III: *Psihologia inconștientului*, „Inconștientul”, traducere din limba germană de G. Lepădatu. București: Trei, e-book].
16. De exemplu, în *The Quest for Consciousness*, Francis Crick și Christoph Koch au explicat că afectul implică „aspectele mai dificile ale conștiinței” (Crick, 2004, p. xiv); așa că ei au „ignorat în mod intenționat [...] modul în care emoțiile [...] contribuie la formarea și modelarea coalițiilor neuronale care sunt suficiente pentru percepția conștientă”, axându-se în schimb pe „aspecte ale conștiinței mai abordabile din punct de vedere experimental”, precum percepția vizuală (Koch, 2004, p. 94). E uimitor că au considerat că mecanismul neocortical al vederii conștiente este o problemă mai simplă decât mecanismul trunchiului cerebral primitiv al afectului, deși au recunoscut în același

timp că afectul modelează coaliţiile neuronale care sunt (aşa-zis) suficiente pentru percepţia conştientă. În orice caz, chiar dacă problema afectului este *într-adevăr* „mai dificilă”, acesta nu este un motiv suficient pentru a o lăsa în afara descrierii pe care o facem conştiinţei. Cam cum ar fi să nu includem efectele cuantice într-o descriere a fizicii!

17. Nu vreau să sugerez că le-au neglijat cu totul; vezi Hume (1748), *Cercetare asupra intelectului omenesc*.
18. Acest termen îi poate fi atribuit lui Panksepp (2011).
19. Skinner (1953), p. 160.
20. Thorndike (1911).
21. Mazur (2013).
22. Vezi Leng (2018), capitolele 16-19, pentru o descriere fascinantă a cercetărilor relevante.
23. Vezi Solms şi Panksepp (2010).
24. Panksepp (1974).

Capitolul 5. Sentimentele

1. Alboni (2014).
2. Următorul paragraf îl parafrazează pe Merker (2007).
3. Vezi Nummenmaa *et al.* (2018) – un articol foarte interesant –, care susţine mai multe dintre concluziile la care ajung eu mai jos, de exemplu, cu privire la imposibilitatea existenţei unor afecte „neutre”, la concretizarea afectelor şi la natura categorială a afectelor.
4. De exemplu, frica aduce cu sine respiraţia rapidă, creşterea ritmului cardiac, redirecţionarea sângelui din intestine spre musculatura scheletului şi, astfel, starea de alertă tonică şi

tendinţa de a fugi. Dacă sentimentul de frică ar aduce cu sine respirația lentă, scăderea ritmului cardiac, redirecționarea sângelui de la musculatura scheletului spre intestine și, astfel, un repaus inert, atunci *nu ar fi frică*. Caracterul intrinsec al afectelor este evident în termografiile corporale totale în prezența diferitelor emoții (vezi Nummenmaa *et al.*, 2018). Vezi și Niedenthal (2007).

5. Discut mai mult despre aceste probleme filosofice în capitolul 11.
6. Urologii numesc această situație *latch-key urgency* („urgență de tip cheia de la casă”).
7. Ekman *et al.* (1987).
8. Barrett (2017).
9. Vezi Panksepp (1998) și Panksepp și Biven (2012) pentru detaliile empirice și referințele bibliografice.
10. Vezi LeVay (1993).
11. Pentru cei interesați de detaliile anatomice: de obicei, la indivizii de sex masculin, centrul circuitului DORINȚEI SEXUALE este hipotalamusul anterior (mai ales nucleii interstițiali), de unde coboară prin nucleii pat ai striei terminale spre PAG. Din punct de vedere chimic, hormonul steroid testosteron (eliberat de testicule și care acționează în mare măsură asupra hipotalamusului anterior) mediază eliberarea în creier a unei peptide numite vasopresină, care explică excitarea sexuală masculină. De obicei, la indivizii de sex feminin, controlul sexual se localizează în hipotalamusul ventromedial, iar substanțele principale sunt estrogenul și progesteronul (amândouă eliberate de ovare, echivalentul la indivizii de sex feminin al testiculelor). Acești hormoni mediază în creier activitatea oxitocinei, o peptidă

care guvernează o bună parte din răspunsul sexual tipic la indivizii de sex feminin. DORINȚA SEXUALĂ este mediată și de alte peptide, cum ar fi LH-RH și CCK.

12. CĂUTAREA este mediată și de neurotransmițătorul numit glutamat și de o serie de peptide, precum oxitocina, neurotensina și orexina.
13. Revin la tema gândirii, cu unele detalii, în capitolul 10.
14. Își are originile în părțile mediale ale amigdalei și trece prin nucleii pat ai striei terminale și prin hipotalamusul median și perifornical, în drum spre PAG. Neuromodulatorul său de comandă este o peptidă numită substanța P, care acționează împreună cu glutamatul și acetilcolina. Acest din urmă fapt ar putea explica de ce tulburarea de comportament REM este atât de frecvent exprimată prin FURIE.
15. Circuitul FRICII apare din amigdala centrală și cea bazolaterală. Din punct de vedere chimic, este mediată de neurotransmițătorul glutamat, pe lângă peptidele DBI, CRF, CCK, α-MSH și NPY.
16. Vezi Tranel *et al.* (2006), care se axează explicit pe emotivitatea *subiectivă* a lui S.M.
17. Vezi Blake *et al.* (2019).
18. Același lucru se aplică și altor sarcini similare, precum „intuiția” în sarcina jocurilor de noroc din Iowa (Turnbull *et al.*, 2014). După cum i-am explicat lui Nicholas Humphrey, care a obiectat la faptul că am folosit termenul „instinct” în loc de „ghicire” în cazul vederii oarbe, totul depinde de întrebările pe care le pui (conferința Science of Consciousness, Interlaken, iunie 2019).
19. LeDoux (1996).
20. Bowlby (1969).

21. De exemplu, Yovell *et al.* (2016) și Coenen *et al.* (2019).
22. Solms și Panksepp (2010).
23. Prin nucleii pat ai striei terminale, aria preoptică și talamusul dorsomedial. Pe lângă mecanismele opioidelor descrise în text, PANICA este mediată de neurotransmițătorul glutamat și de neuropeptidele oxitocină, prolactină și CRF.
24. Acesta este probabil motivul pentru care durerea se somatizează atât de frecvent ca durere fizică. Vezi Eisenberger (2012) și Tossani (2013).
25. GRIJA este mediată și de dopamină.
26. Circuitul GRIJII coboară din cortexul cingular anterior, prin nucleii pat ai striei terminale, aria preoptică și aria tegmentală ventrală până la PAG.
27. Forrester *et al.* (2018).
28. Penksepp și Burgdorf (2003). Vezi www.youtube.com/watch?v= j-admRGFVNM
29. Sper că acesta nu mai este un joc popular. Nu *aprob* faptul că amerindienii au fost dominați de fermieri coloniști, dar cu siguranță au fost.
30. Empatia derivă din atitudinea intențională sau „teoria minții" care, desigur, nu este la fel de dezvoltată la toate speciile de mamifere. Prin urmare, dezvoltarea empatiei nu e în niciun caz un proces automat, după cum ar putea sugera teoria „neuronilor-oglindă". Empatia nu este un reflex, ci o realizare evolutivă (vezi Solms, 2017a).
31. Totuși, talamusul dorsomedial și aria parafasciculară par să fie deosebit de importante, ca și PAG, desigur. Dacă există neurotransmițători principali ai circuitului JOACĂ, atunci vorbim de opioidele mu, dar asta ar putea pur și simplu să

reflecte faptul că siguranța (adică nivelul scăzut de PANICĂ/MÂHNIRE) este o precondiție necesară pentru JOACĂ. Alți posibili modulatori ai circuitului JOACĂ sunt glutamatul, acetilcolina și canabinoidele. Circuitul talamocortical identificat de Zhou și colaboratorii săi (2017) are legătură numai cu un aspect specific al circuitului JOACĂ, şi anume dominarea. Vezi și Van der Westhuizen și Solms (2015), Van der Westhuizen *et al.* (2017).

32. Pellis și Pellis (2009).

Capitolul 6. Sursa

1. Moruzzi și Magoun (1949).
2. Fischer *et al.* (2006). Leziunea era în apropierea (chiar deasupra) nucleului parabrahial medial. E interesant că în această regiune Hobson a identificat celulele-sursă colinergice pentru somnul REM. Vezi și Parvizi și Damasio (2003), Golaszewski (2016).
3. Blomstedt *et al.* (2008). Electrodul a fost introdus în substanța neagră. Locul vizat era nucleul subtalamic.
4. Damasio *et al.* (2000), Holstege *et al.* (2003).
5. Mă refer la convexitatea neocortexului, nu la hipocamp.
6. Garcia-Rill (2017).
7. Holeckova *et al.* (2006).
8. Această utilizare a termenilor „canal” și „stare” i se datorează lui Mesulam (2000). În capitolul 9 folosesc analogia „modurilor de operare” pentru funcțiile „stărilor”. Pentru o descriere interesantă a aspectului concret al creierului în diferite stări sau moduri de operare, la nivel celular, vezi

Abbott (2020).

9. Panksepp (1998), p. 314. Acest citat din Panksepp continuă mai jos și include următoarele două note de final.

10. Bailey și Davis (1942).

11. Depaulis și Bendler (1991).

12. Vezi Walker (2017), pentru o prezentare foarte bună a modului în care funcționează acest ritm. Chiar și moluștele și echinodermele (precum stelele-de-mare) prezintă un ciclu somn/veghe. Panksepp (1998), p. 135, arată că reglarea somnului este mai veche din punct de vedere filogenetic decât sistemul reticulat activator. Pornind de la această idee, el face o sugestie contrariantă: „Ceea ce acum este mecanismul somnului REM la origini media activarea selectivă a emotivității. Înainte de apariția unor strategii cognitive complexe, e posibil ca animalele să-și fi generat cea mai mare parte a comportamentului din rutine psihocomportamentale al proceselor primare pe care acum le recunoaștem drept sisteme emoționale primitive [...]. Cu alte cuvinte, e posibil ca multe dintre comportamentele animalelor preistorice să fi apărut în mare parte din subrutine emoționale preprogramate. Aceste soluții comportamentale simpliste au fost înlocuite în cele din urmă de abordări cognitive mai sofisticate care au necesitat nu numai mai mult neocortex, ci și mecanisme de activare noi, pentru a susține funcții eficiente în starea de veghe în interiorul acelor zone cerebrale emergente”.

În lumina a ceea ce propun mai jos, sugestia lui Panksepp poate fi reformulată astfel: prozencefalul adaugă la programele motoare instinctuale automate, inferioare o capacitate de a modula *din punct de vedere contextual* comportamentul

emoțional – și, astfel, de a învăța din experiență.

13. Dar nu vorbim despre conceptul filosofic de intenționalitate sau despre „pe cine vizează intenția”. Panksepp are în vedere ceva precum „volițiunea”. Totuși, când mă voi referi la conceptul filosofic, mai târziu, va deveni evident că e profund legat de volițiune.

14. Astfel, în analiza exhaustivă făcută de Pfaff (2005) pe tema activării, pe care o descrie drept „forța fundamentală din sistemul nervos”, autorul operaționalizează termenul după cum urmează: „«Activarea generalizată» este mai intensă la un animal sau la un om care este: (S) mai receptiv la stimuli senzoriali de toate tipurile, (M) mai activ sub aspect motor și (E) mai reactiv din punct de vedere emoțional”. Având în vedere că „activarea” este un element central în cartea de față, reproduc un fragment mai lung din Pfaff (2005) în anexa de la p. 313, care oferă și o punte utilă spre subiectul capitolului următor. Sunt recunoscător că am această ocazie să subliniez importantele cercetări ale lui Donald Pfaff, care (deja la începutul anilor 1990, când l-am cunoscut) a manifestat o apreciere neobișnuită pentru ideea de „pulsiune” a lui Freud.

15. PAG are proiecții spre toți nucleii sursă a neuromodulatorilor din sistemul reticulat activator. Celelalte destinații importante din trunchiul cerebral ale proiecțiilor PAG sunt hipotalamusul medial, nucleul cuneat, formația reticulată pontină, nucleul solitar, nucleul *gracilis*, nucleul dorsoreticular și măduva ventrolaterală. Vezi Linnman *et al.* (2012).

16. Venkatraman, Edlow și Immordino-Yang (2017). Nu îmi place cuvântul „descendent” în acest context, deoarece PAG

integrează feedback afectiv cerebral superior și visceral inferior. Ea „coboară” doar în sensul că duce la un output motor. „Centripet” ar fi un cuvânt mai potrivit, care ar putea fi pus în contrast cu o rețea „centrifugă” (deci cu ceea ce Edlow numește rețea „modulatoare”). O rețea „centripetă” ar include rețele atât „descendente”, cât și „ascendente”. Linnman și colaboratorii săi (2012) descriu în consecință PAG drept locul *interacțiunii dintre* sistemul „limbic descendent” și cel „senzorial ascendent”.

17. Venkatraman, Edlow și Immordino-Yang (2017).
18. *Ibidem*.
19. Linnman *et al.* (2012), p. 517, sublinierea mea.
20. *Ibidem*, sublinierea mea. Deloc surprinzător, în studiile de imagistică cerebrală umană, PAG pare să aparțină unei rețele „proeminente” (Seeley *et al.*, 2007).
21. Ezra *et al.* (2015), p. 3468, sublinierea mea.
22. Panksepp și Biven (2012), p. 413, sublinierea mea.
23. Linnman *et al.* (2012), p. 506, sublinierea mea.
24. Panksepp și Biven (2012).
25. Se obișnuia să fie numită „griul central”.
26. Cele „din spate” sunt PAG laterală și PAG dorsolaterală. Cea „din față” este PAG ventrolaterală. Această clasificare neglijează PAG dorsomediană.
27. Vezi Venkatraman, Edlow și Immordino-Yang (2017): „Când este stimulată, această coloană produce vocalizare emoțională, confruntare, agresiune și activare simpatică, manifestate printr-o creștere a tensiunii arteriale, a ritmului cardiac și a celui respirator [...]. În interiorul acestei coloane dorsolaterale/laterale există două părți. Partea rostrală este responsabilă pentru putere/dominare (generând un răspuns

de luptă), iar partea caudală invocă frica (generând un răspuns de fugă) cu creșterea fluxului sangvin spre membre”.

28. Coloana din față „primește semnale de durere somatice și viscerale «lente, arzătoare» slab localizate și, dacă este stimulată, produce *coping* pasiv, comportament de bolnav pe termen lung, încremenire cu hiporeactivitate și o inhibare a refluxului simpatic [...]. În acest fel, probabil că este implicată în emoțiile de fond precum cele care contribuie la dispoziție” (*ibidem*).
29. Simone Motta și colegii săi s-au exprimat astfel (Motta, Carobrez și Canteras, 2017, p. 39): „[PAG] este considerată în general un loc în aval în rețelele neuronale pentru exprimarea unor comportamente diverse, generând răspunsuri stereotipe. Dar tot mai multe dovezi sugerează că PAG poate să exercite o modulare mai complexă a multor răspunsuri comportamentale și că poate să funcționeze ca un centru unic de furnizare a unui ton emoțional primar pentru a influența pozițiile prozencefalice care mediază răspunsuri complexe de aversiune și de năzuință”.
30. În PAG are loc frecvent stimularea cerebrală profundă pentru tratamentul durerii cronice, dar nu reduce capacitățile somato-senzoriale corticale.
31. Merker (2007). La rândul lui, Merker le atribuie această perspectivă lui Penfield și Jasper (1954).
32. Pentru scopurile acestei cărți, am simplificat termenul tehnic „triunghi de selectare mezodiencefalic”. De asemenea, modul în care folosesc termenul lui Merker se referă mai mult la o *interfață* în decizie decât la un *triunghi* (adică o interfață între nevoie și context). După cum vom

vedea în capitolul următor, acțiunea („selectarea acțiunii” pentru Merker) și percepția („selectarea țintei” pentru Merker) – care împreună alcătuiesc contextul – sunt fețele aceleiași medalii. Iată cum descrie Merker (2007), p. 70, acest lucru: „Oricât s-a extins ulterior telencefalul, chiar până la a îngropa mezodiencefalul sub un neocortex de mamifer ce s-a dezvoltat rapid, nu a fost niciodată nevoie de vreun alt aranjament, iar asta din motive cu adevărat fundamentale. Niciun nerv eferent nu își are nucleul motor situat mai sus de etajul mezencefalului. Asta înseamnă că prin intersecția foarte îngustă din trunchiul cerebral dintre mezencefal și diencefal [...] trec toate informațiile prin care prozencefalul poate vreodată să genereze, să controleze sau să influențeze comportamentul de orice tip”. Merker numește această intersecție „gâtuirea sinencefalică”. Și adaugă: „Nu e nevoie să știm mai mult despre creierul vertebratelor decât faptul că motoneuronii rostrali sunt localizați *sub* gâtuirea sinencefalică, ca să știm că conținutul informațional total al prozencefalului trebuie să treacă printr-o reducere masivă a datelor în cursul translatării sale în timp real în comportament”. Faptul că decizia privind ceea ce e de făcut în continuare se ia la acest nivel (al trunchiului cerebral) – adică după ce regiunile prozencefalului și-au prezentat „cererile” – este ilustrat în mod dramatic de exemplul alarmei de sufocare, discutat mai sus: toate considerațiile cognitive sunt *trecute într-un plan secundar* de senzația de foame de aer, care se declanșează la nivelul trunchiului cerebral. Centrii controlului respirator sunt localizaţi în punte și în bulbul rahidian.

33. Panksepp (1998), p. 312, a descris aranjamentul funcțional al

SINELUI, cu un deceniu înaintea lui Merker, după cum urmează: „Straturile mai profunde ale coliculilor constituie o reprezentare motoare elementară a corpului [obiectiv], nu care interacționează numai cu sistemele vizual, auditiv, vestibular și somatosenzorial, ci și cu circuitele emoționale învecinate din PAG. PAG elaborează o hartă diferită, de tip visceral a corpului [subiectiv] alături de reprezentările neuronale de bază ale sistemelor durerii, fricii, furiei, detresei de separare, sexuale și de comportament matern (după cum este rezumat [în capitolul anterior] în această carte). Lângă PAG se află regiunea mezencefalică locomotoare, ce este capabilă să declanșeze tipare neuronale care ar trebui să fie un substrat esențial pentru a iniția diferite tendințe de acțiune coerentă". De asemenea, Damasio și Carvalho (2013) au propus acest aranjament funcțional general pentru ceea ce au numit „protosine". Au arătat că alte hărți senzoriale heteromodale primitive ale corpului sunt furnizate de complexul parabrahial și de nucleul tractului solitar. Aceștia ar putea fi precursorii evolutivi ai funcției integratoare îndeplinite de triunghiul de decizie din mezencefal.

34. Următorul citat este din Merker (2007), p. 73: „[Sinele conștient] este singur și localizat în spatele rădăcinii nasului, în interiorul capului. De acolo *se pare* că ne confruntăm cu lumea vizibilă direct printr-o singură deschizătură ciclopică goală aflată în fruntea capului (Hering, 1879; Julesz, 1971). Totuși, în mod evident, e o simplă aparență, întrucât, dacă am fi literalmente și în fapt situați în interiorul capului propriu, ar trebui să vedem nu lumea, ci țesuturile anatomice din craniu atunci când privim. Deschizătura

ciclopică este o invenție neuronală convenabilă prin care lumea vizuală distală este «inserată» printr-o parte lipsă a corpului vizual proxim, care este «fără cap», cum s-ar spune, sau, mai precis, căruia îi lipsește partea superioară a feței (vezi Harding, 1961). În schimb, somestezia menține o continuitate nealterată în această regiune. Deschizătura goală prin care privim lumea trădează natura simulată a corpului și a lumii care ne sunt date în conștiință”.

35. *Ibidem*, p. 72. Vezi Stoerig și Barth (2001) pentru o simulare plauzibilă. Ne facem astfel o idee despre lumea senzorimotoare probabilă a copilului cu hidroanencefalie și a animalului care a suferit o decorticare cerebrală (*cf.* și vederea oarbă).
36. White *et al.* (2017), Panksepp (1998), p. 311.
37. Merker (2007), p. 72.
38. Hohwy (2013).
39. După cum s-a și întâmplat, la câțiva ani de la intervenția chirurgicală, din cauza unei infecții a căilor respiratorii superioare (altfel ușor tratabilă) care a fost diagnosticată prea târziu.
40. Vezi Friston (2005).
41. Hohwy (2013), Clark (2015).
42. Să ne gândim, de exemplu, la neuronii mediali dorsali pereche ai musculiței de oțet, *Drosophila*. În mod surprinzător, pare chiar să fi existat niște precursori primitivi la nematodul *Caenorhabditis elegans* (vezi Bentley *et al.*, 2016; Chew *et al.*, 2018).
43. Următoarea imagine din Freud (1925), p. 231, ne-ar putea ajuta să ne imaginăm aranjamentul funcțional descris și, în același timp, ne-ar putea permite să-i înlocuim termenii

„metapsihologici” cu unii fiziologici: „Inervațiile de investiție sunt trimise și din nou retrase din interior spre sistemul *P-Cs* [sistemul „conștiinței perceptive”], prin șocuri periodice rapide. Atât timp cât sistemul este în acest mod investit, el primește percepții însoțite de conștiință și conduce în continuare excitația în sistemele mnezice inconștiente; îndată ce investirea este retrasă, conștiința se stinge, activitatea sistemului fiind sistată. Ar fi ca și cum [sinele] ar întinde, prin intermediului sistemului *P-Cs*, o antenă în lumea exterioară, care apoi este repede retrasă, îndată ce i-a gustat excitațiile” [Sigmund Freud, *Opere esențiale*, vol. III: *Psihologia inconștientului,* „Notă asupra «blocnotesului» magic”, traducere de G. Lepădatu. București: Trei, e-book]. „Investiția” este o activare modulatoare. Prin urmare, „inervațiile de investiție” care pipăie percepția corticală în această imagine figurativă sunt impulsuri ale activării nucleului cerebral. Am înlocuit termenul „inconștient” al lui Freud cu „sinele” în acest citat ca să evit faptul că el a combinat în mod eronat aceste două sisteme (Solms, 2013).

44. Merker (2007). Panksepp și Biven (2012), pp. 404-405, folosesc ritmul teta al hipocampului ca ilustrare a modului în care această „monedă comună” ar putea să funcționeze din punct de vedere fiziologic, având în vedere faptul că hipocampul codifică *contextul*: „Există indicii sugestive în literatura tradițională din neuroștiințe ale anumitor tipuri de oscilații sincrone relevante din creier, precum ritmurile de 4-7 Hz din hipocamp numite *ritm teta*, care ajută animalele să exploreze lumea (de exemplu, adulmecatul la șobolani), creând astfel amintiri în hipocamp. Ritmul teta este semnătura neuronală foarte caracteristică a

hipocampului atunci când prelucrează activ informații. Acest ritm este evident mai ales în timpul activării artificiale a sistemului CĂUTĂRII la șobolani, un sistem emoțional de primă importanță pentru culegerea informațiilor când animalele adulmecă și își explorează mediul (Vertes și Kocsis, 1997). Cu alte cuvinte, ritmul adulmecării corespunde de obicei frecvenței curente a undelor teta din hipocamp [...]. Acest lucru ar putea evidenția modul în care cunoașterea cognitivă apare din activările modelate ale proceselor afective".

45. „Pipăit" este termenul colocvial pentru senzația somatică, ce conține submodalități precum propriocepția, detectarea temperaturii, detectarea vibrației – ca și alte modalități perceptive.
46. Când sunt predominante.
47. Solms (2013).
48. Se poate spune că acetilcolina modulează încrederea în semnalele de eroare, dar această afirmație este o simplificare exagerată. Generalizări la fel de exagerat de mari se pot face şi cu privire la serotonină în legătură cu semnalele predictive, la dopamină în legătură cu stările active și la noradrenalină în legătură cu stările senzoriale. Vezi Parr și Friston (2018), pentru o opinie mai elaborată.
49. „Fizică" (φυσική) înseamnă „cunoașterea naturii" – a întregii naturi, nu numai a lucrurilor pe care poți să le vezi și să le atingi. Ea oferă explicații fundamentale ale fenomenelor naturale. Mulți presupun că fizica studiază doar materia – excluzând astfel mintea prin definiție –, dar asta ar însemna că mintea nu face parte din natură, ceea ce nu poate să nu ducă la însăși tema cărții de față. „Un

fenomen natural e complet explicat din punct de vedere fizic numai după ce i s-au descoperit originile până la forțele supreme ale naturii în care se întemeiază și care sunt active în el.” (Helmholtz, 1892) Se dovedește că materia este o stare a energiei (de unde și $E = MC^2$). „Forțele primare” *explică* fenomenele superficiale, ele nu sunt observate direct, ci sunt deduse. Din acest motiv, ele sunt raportate științific în termeni nonfenomenologici, ca abstracțiuni.

Cf. Critica rațiunii pure a lui Kant: „Deci numai prin faptul că supunem succesiunea fenomenelor, prin urmare orice schimbare, legii cauzalității este posibilă experiența însăși, adică o cunoaștere empirică a acestor fenomene; prin urmare, ele însele nu sunt posibile, ca obiecte ale experienței, decât după această lege” [Immanuel Kant, *Critica rațiunii pure*, ed. a III-a îngrijită de Ilie Pârvu, traducere de Nicolae Bagdasar și Elena Moisuc. București: IRI, 1998, p. 113]. Abstracțiunile matematice sunt preferabile din punct de vedere convențional celor verbale, întrucât cer ca forțele deduse (și relațiile dintre ele) să fie măsurabile și cuantificabile. Acest lucru furnizează o monedă numerică comună în spatele diverselor aparențe fenomenologice prin care pot fi calculate relațiile legitime dintre „forțele supreme” ale naturii. După cum a spus Galileo Galilei: „Cartea naturii este scrisă în limbajul matematicii”.

Capitolul 7. Principiul energiei libere

1. Cristalele își minimizează energia liberă într-un mod simplu, deoarece starea lor constantă de neechilibru tinde spre un

punct fix. Adică ele doar se aranjează în tipare compacte și rămân așa, chiar și atunci când sunt ușor perturbate. Situația devine mai complexă atunci când elementul de atracție are o structură itinerantă cu o dinamică de tipul celei căreia i se conformează creierul uman.

2. Pe scurt, indicele Hirsch este numărul lucrărilor publicate de cineva care sunt citate (de egalii săi) de mai multe ori decât rangul lor în succesiunea lucrărilor sale publicate citate. Astfel, dacă a patruzecea cea mai citată lucrare publicată a sa este citată de 42 de ori, dar a patruzeci și una cea mai citată lucrare publicată este citată de doar 39 de ori (adică mai puțin de 41), atunci indicele Hirsch al autorului este 40.
3. La 20 iulie 2020.
4. Carhart-Harris și Friston (2010).
5. Freud (1894), p. 60.
6. Friston (2005), sublinierea mea. Această afirmație e valabilă chiar și în cazul articolului său scris împreună cu Carhart-Harris, care începe așa (Carhart-Harris și Friston, 2010, p. 1265, sublinierea mea): „Descrierile făcute de Freud proceselor primare și secundare corespund activității autoorganizate din sistemele *corticale* ierarhice și [...] descrierile pe care le face Eului corespund funcțiilor modului implicit, iar schimburile reciproce sistemelor cerebrale subordonate. Această descriere neurobiologică se bazează pe o perspectivă asupra creierului ca inferență ierarhică sau ca aparat Helmholtz. Din această perspectivă, rețele intrinsece la scară mare ocupă niveluri supraordonate ale sistemelor cerebrale ierarhice care încearcă să-și optimizeze reprezentarea întregului sistem senzorial. Această optimizare este formulată ca minimizând o energie

liberă, un proces similar din punct de vedere formal cu abordarea energiei în formulările freudiene”.

7. Friston (2013).

8. Solms și Friston (2018).

9. Parvizi și Damasio (2003).

10. Vezi Damasio (2018) și recenzia mea la această lucrare, Solms (2018a).

11. Ideea lui Einstein la care mă refer este natura cuantică a luminii (Einstein, 1905). Stephen Hawkins a replicat: „Nu numai că Dumnezeu e clar că joacă zaruri, ci uneori ne şi bulversează aruncându-le unde nu pot fi văzute. Mulți oameni de știință sunt ca Einstein, în sensul că sunt profund atașați emoțional de determinism. Spre deosebire de Einstein, ei acceptă reducerea capacității noastre de a prezice, reducere ce vine din teoria cuantică”. Se dovedește și că fenomenele conștiinței sunt predictibile doar *la un nivel probabilistic*.

12. Asta se întâmpla la University of Southern California, în aprilie 2018. Într-o convorbire telefonică din ianuarie 2019, după ce a citit versiunea publicată a articolului nostru, încă apăra „știința noastră” pe baza căreia conștiința era intrinsec *biologică*. Totuși, din fericire, și-a schimbat opinia de atunci; vezi Man și Damasio (2019).

13. E interesant că Hermann von Helmholtz (unul dintre elevii lui Johannes Müller care a înființat Societatea de Fizică din Berlin) a jucat un rol important în formularea acestei legi.

14. Această afirmație este valabilă în realitate, în limitele termodinamice, dar nu (strict vorbind) și în teorie. Din punct de vedere teoretic, ar fi mai corect să spunem: conform celei de-a doua legi, este foarte, foarte *improbabil* ca procesele

naturale să fie reversibile. Asta deoarece, când rezolvi ecuațiile din orice sistem, ai nevoie nu numai de legile relevante ale dinamicii, ci și de *condițiile inițiale* ale sistemului – iar acestea strică simetria ecuațiilor fundamentale. Prin urmare, dacă am putea începe cu toate bucățile unei cești sparte mișcându-se una spre alta la viteza corectă și toate undele sonore mișcându-se înapoi spre ceașcă exact așa cum trebuie și la fel și energia care a trecut pe podea când a aterizat ceașca, atunci totul ar putea reveni ca să recreeze ceașca inițială intactă. Totuși, pentru că nu suntem niciodată în această situație, nu vedem niciodată entropia reducându-se în realitate.

15. Cu excepția unor perioade foarte, foarte scurte.

16. Când la o substanță se adaugă căldură, moleculele și atomii vibrează mai repede. Pe măsură ce atomii vibrează mai repede, spațiul dintre ei crește. Mișcarea și spațiul dintre particule determină starea materiei din substanță. Rezultatul final al mișcării moleculare sporite este că substanța se dilată și ocupă mai mult spațiu.

17. Din punct de vedere tehnic, entropia este asociată cu energia risipită, dar nu e același lucru. Entropia nu are dimensiuni, spre deosebire de energie.

18. Din punct de vedere tehnic, entropia e asociată cu numărul de stări echivalente ce corespund unei configurații macroscopice date.

19. Fizicianul Alan Lightman (2018), pp. 67-68, formulează excelent: „Dacă divizăm neobosit spațiul în părți din ce în ce mai mici, așa cum a făcut Zenon, căutând cel mai mic element al realității, odată ce ajungem la lumea fantasmagorică a lui Planck, spațiul nu mai are semnificație.

Cel puțin, ceea ce *înțelegem* prin «spațiu» nu mai are semnificație. În loc să răspundem la întrebarea care este cea mai mică unitate a materiei, invalidăm cuvintele folosite pentru a pune întrebarea. Poate că aceasta este calea întregii realități supreme, dacă există așa ceva. Pe măsură ce ne apropiem, pierdem vocabularul”.

Un alt fizician, Carlo Rovelli (2014), p. 167, face o descriere mai prozaică: „Fundalul spațiului dispare, timpul dispare, particulele clasice dispar, împreună cu câmpurile clasice. Atunci din ce este făcută lumea? Acum răspunsul e simplu: [...] lumea este făcută în întregime din câmpuri cuantice”. Chiar și incertitudinea cuantică face parte din universul *fizic*.

20. Un „byte” conține opt biți. Prin urmare, un „gigabyte” are opt miliarde de biți. *Viteza* prelucrării informațiilor este exprimată în „gigahertzi”, un GHz fiind un miliard de biți pe secundă. Realitatea fizică a informației este reflectată de faptul că aceste unități sunt şi măsurabile, și cumpărabile (de exemplu, de la furnizori de servicii de internet).

21. Vezi anexa de la p. 313. Aici trebuie să definim cine este nesigur de ce anume. Aceeași afirmație este valabilă şi în cazul cuvântului „comunică” din fraza de după cea care conține nota 22: cine comunică cu cine? Te rog să citești mai departe.

22. Gosseries *et al.* (2011).

23. Dacă crezi că această afirmație combină entropia informațională cu entropia termodinamică, vezi citatul din Tozzi, Zare și Benasich în nota 61, mai jos.

24. După cum vei vedea, e crucial faptul că avem nevoi *multiple*, inclusiv CĂUTAREA.

25. Ia în considerare, de exemplu, inseparabilitatea cuantică dintre două particule: o particulă „conţine informaţii” despre cealaltă.
26. Jaynes (1957). Creșterea volumului sistemului descris în exemplul nostru formal de mai sus (gazul din incintă) îi crește entropia termodinamică deoarece crește numărul de microstări posibile ale sistemului care corespund valorilor măsurabile ale variabilelor macroscopice ale acestuia, astfel făcând mai încărcată cu informații orice descriere completă a stării sale. Mai exact: în cazul distinct ce folosește logaritmi în baza doi, entropia termodinamică redusă este egală cu numărul minim de întrebări binare la care trebuie să se răspundă pentru a specifica complet microstarea, dat fiind că cunoaștem macrostarea. Se poate găsi o relație directă și reală din punct de vedere fizic între entropia termodinamică și entropia informațională dacă se atribuie o unitate de măsură fiecărei microstări ce apare per unitate de substanță omogenă, apoi se calculează entropia termodinamică a acestor unități. Prin analiză teoretică sau prin observație, microstările vor apărea cu probabilități diferite, iar acest lucru va determina entropia informațională. Asta arată că entropia lui Shannon este o unitate de măsurare statistică valabilă a microstărilor ce *nu are o altă unitate fizică fundamentală decât unitățile de informație*.
27. Shannon (1948).
28. De exemplu, „principiul aspectului dublu” formulat de Chalmers, discutat în capitolul 11.
29. Wheeler i-a fost student lui Niels Bohr, care a formulat principiul complementarității. Conform acestui principiu, obiectele au proprietăți complementare ce nu pot fi

observate sau *măsurate* simultan. Exemple de proprietăți complementare sunt particula și unda, poziția și momentul, energia și durata, rotirea pe axe diferite, valoarea unui câmp și schimbarea sa (într-o anumită poziție), precum și inseparabilitatea și coerența.

30. Wheeler a spus „ceva din bit” (*it from bit*). Iată citatul (Wheeler, 1990, p. 5): „Ceva din bit. Cu alte cuvinte, fiecare ceva – fiecare particulă, fiecare câmp de forță, însuși continuumul spațiu-timp – își derivă funcția, semnificația, chiar întreaga existență – chiar dacă indirect în unele contexte – din răspunsurile obținute de mașinării la întrebări de tip da sau nu, opțiuni binare, biți. *Ceva din bit* simbolizează ideea că fiecare element din lumea fizică are la bază – o bază foarte profundă, în majoritatea cazurilor – o sursă și o explicație imateriale”.
31. *Ibidem*. Cuvintele citate aici urmează imediat citatului de mai sus.
32. Cititorii atenți poate observă aici că, în sfârșit, mă ocup de întrebarea pe care mi-am pus-o în copilărie, despre primul răsărit de soare, după cum am discutat în introducere.
33. Darwin (1859). Vezi Friston (2013). Rovelli (2014), pp. 225-226, oferă o explicație lucidă a acestui lucru: „Un organism viu este un sistem ce se reformează continuu, interacționând neîncetat cu lumea exterioară. Dintre asemenea organisme, continuă să existe numai cele care sunt mai eficiente în acest sens și, prin urmare, organismele vii manifestă proprietăți care le dau posibilitatea să supraviețuiască. Din acest motiv, ele sunt interpretabile, iar noi le interpretăm, în termeni de intenționalitate, în funcție de [obiectiv și] scop. Prin urmare, aspectele finalistice ale lumii biologice (aceasta

este descoperirea crucială a lui Darwin) sunt rezultatul selecției unor forme complexe eficiente în a persista. Dar modul eficient de a continua să existi într-un mediu schimbător constă în a gestiona mai bine corelațiile cu lumea exterioară, adică informațiile; în a culege, a stoca, a transmite și a elabora informații. Din acest motiv există ADN, sisteme imunitare, organe senzoriale, sisteme nervoase, creieri complecşi, limbaje, cărți, biblioteca din Alexandria, calculatoare și Wikipedia: ele maximizează eficiența gestionării informațiilor – gestionarea corelațiilor care favorizează supraviețuirea".

34. Aici nu contează că sistemele cu autoorganizare sunt întotdeauna vii (nu sunt), ci că întotdeauna sistemele vii se autoorganizează. Și mai important: nu toate sistemele cu autoorganizare sunt conștiente. Despre fizica originilor vieții, vezi England (2013), pentru o analiză interesantă dintr-o perspectivă similară cu a lui Friston. Teoria „adaptării impulsionate de disipare" formulată de England afirmă că grupurile de atomi care sunt impulsionați de surse externe de energie tind să apeleze la aceste surse, aliniindu-se și rearanjându-se, astfel încât să absoarbă energia mai bine și să o disipeze sub formă de căldură (adică își minimizează entropia în detrimentul mediului). Mai departe, el arată că această tendință de disipare este propice autoreplicării: „O modalitate extraordinară de disipare suplimentară constă în a face mai multe exemplare din tine însuți". Totuși, pentru *biologia* originii vieții, vezi Lane (2015).
35. Ashby (1947). Vezi și Conant și Ashby (1970).
36. Friston (2013), p. 6. Pe monitorul calculatorului se vedeau,

desigur, *reprezentări* ale subsistemelor, care nu trebuie confundate cu dinamica statistică. Această afirmație e valabilă și în cazul qualia perceptive asociate cu funcționarea corticală, după cum vom vedea.

37. Andrei Markov (1856-1922) a fost un matematician rus strălucit, ca și fratele și fiul lui. A lucrat în principal în domeniul proceselor stocastice și este celebru astăzi pentru noţiunile de lanțuri Markov și procese Markov. A fost rebelul perfect. Descurajarea instituțională a fost atât de puternică, încât nu a primit recunoașterea academică în timpul vieții. În 1912, a protestat împotriva excomunicării lui Lev Tolstoi din Biserica Ortodoxă Rusă, cerând să fie şi excomunicat. Biserica a dat curs cererii sale.

38. Prin „invers” nu vreau să spun că stările interne ale sistemului sunt ascunse de stările externe (nu-sistem), ci că *punctul de vedere* al sistemului este ascuns; e disponibil doar pentru sine. Când e vorba de problema altor minți, un sistem nu poate cunoaște niciodată stările interne ale altui sistem nu numai din cauză că toate stările externe sunt ascunse de el, ci și din cauză că stările interne ale altor sisteme sunt interne exclusiv pentru acele sisteme.

39. Friston (2013), p. 8.

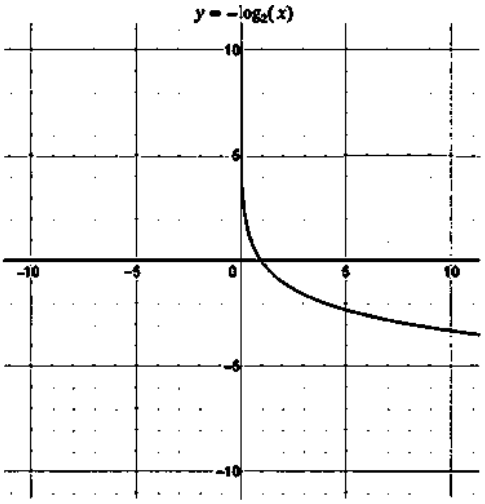
40. Termenul „autopoiezǎ” a fost introdus de Maturana și Varela (1972) pentru a defini chimia autonomă a celulelor vii.

41. Având în vedere ce am spus mai sus despre capacitatea de reprezentare a sistemului, cititorii amatori de filosofie vor observa că determină și „intenționalitatea” în sensul lui Brentano (1874).

42. Friston (2013), p. 2.

43. Sau, pur și simplu, este ascunsă și indisponibilă pentru lucru mecanic util.
44. Este ceea ce Helmholtz a numit energie care nu este liberă (*TS* în ecuație). Distincția dintre energia „liberă” și cea „legată” cu siguranță va fi apreciată de specialiștii în opera lui Freud. Vezi capitolul 10, nota 16, despre noțiunea de „proces secundar” în accepțiunea lui Freud.
45. $A = U + pV - TS$, unde p este presiunea și V volumul. Această ecuație cuantifică energia liberă a sistemelor a căror activitate este asociată cu dilatarea și compresia sistemului la o temperatură și o presiune constante.
46. Ecuația are o formă lungă și una scurtă. Iat-o pe cea lungă: $F(s, \mu) = Eq[-\log p(s, \psi \mid m)] - H[q(\psi \mid \mu)]$, unde F (energie liberă „variațională” sau, pe scurt, energie liberă Friston) este izomorfa energiei libere Gibbs și a energiei libere Helmholtz. În ce privește celelalte cantități din ecuație, s denotă stările senzoriale (ale zonei Markov, discutată mai sus), μ stările interne, Eq energia medie, $p(s, \psi \mid m)$ o densitate de probabilitate cu privire la stările senzoriale și externe (ascunse) într-un model generativ m, H denotă entropia, iar $q(\psi \mid \mu)$ o densitate variațională cu privire la stările ascunse parametrizate de stări interne. Relația dintre această ecuație folosită în știința informației și cea folosită în termodinamică nu este imediat evidentă. Totuși, când ecuația lungă este comprimată după cum urmează, ea seamănă cu cea din termodinamică: $F = Eq - H$. Aici, F denotă energia liberă Friston, Eq energia medie, iar H entropia.
47. Adică informația medie obținută după multe măsurători ale microstărilor.

48. Practic, expresiile ulterioare citate mai sus (referitoare la surprindere și divergență) ne spun că, deși energia liberă Helmholtz arată energia disponibilă pentru acțiune efectivă, energia liberă Friston arată *diferența dintre modul în care lumea* respectivă *este modelată de un sistem și modul în care lumea se comportă de fapt*. (Voi explica imediat ce legătură are acest lucru cu acțiunea.)
49. Complexitatea termenului „energie liberă Friston” are același punct fix ca energia liberă Helmholtz (presupunându-se că sistemul este închis din punct de vedere termodinamic, dar nu izolat). Dacă perturbațiile senzoriale sunt suspendate pe o perioadă corespunzător de lungă, complexitatea este minimizată deoarece acuratețea poate fi neglijată. În acest punct, sistemul este în echilibru, iar stările interne minimizează energia liberă Helmholtz conform principiului energiei minime (care este, practic, o reformulare a legii a doua).
50. Enunț formal: surprinderea este logaritmul negativ al probabilității rezultatului (s), într-un model dat al stărilor ascunse ale lumii ([$-\log p(s, \psi \mid m)$] în ecuația de mai sus). Pe măsură ce probabilitatea (axa x) se îndreaptă spre 0, surprinderea (axa y) urcă; pe măsură ce probabilitatea se îndreaptă spre 10, surprinderea coboară.



51. Astfel, minimizarea surprinderii despre lucrurile care s-au întâmplat va minimiza, în medie, entropia. (Surprinderea este un atribut al datelor sau al observațiilor, iar energia liberă este un atribut al credințelor. Prin urmare, partea entropică din energia liberă nu este surprinderea medie a observațiilor, ci entropia credințelor despre cauzele latente ale observațiilor.)
52. Energia liberă Friston (energia medie minus entropia) este echivalentă cu surprinderea (exprimată ca –log $p(s \mid m)$) plus „divergența perceptivă” (exprimată ca $DKL\ [q(\psi \mid \mu) \mid\mid p(\psi \mid s, m)]$), care este întotdeauna mai mare decât sau egală numai cu surprinderea. În esență, „surprinderea” medie înseamnă entropie (informațională), după cum explic mai jos. „Divergența perceptivă” măsoară diferența dintre evenimentele presupuse și evenimentele reale, într-un model generativ. *DKL* denotă divergența perceptivă. Vine de la divergența Kullback-Leibler, numită și entropie relativă, care cuantifică divergența dintre două densități de probabilitate: cele două care ne interesează sunt despre stările ascunse – densitatea variațională codificată în stările interne (de exemplu, neuronale) și densitatea condițională

reală, date fiind stări senzoriale. *DKL* este întotdeauna mai mare decât sau egală cu zero. Intuim că asta se datorează faptului că logaritmii negativi (explicați mai jos) au întotdeauna reprezentări grafice aproximativ în formă de U, așa că o linie ce unește două puncte pe U nu poate fi niciodată sub partea inferioară a U-ului (termenul tehnic este „funcție convexă”). Astfel se asigură că energia liberă impune surprinderii o limită în partea superioară.

53. Vezi Clark (2017): „Autovalidarea [...] apare atunci când o ipoteză explică cel mai bine unele dovezi și, în virtutea acestui succes explicativ, aduce astfel dovezi în sprijinul propriului adevăr sau al propriei corectitudini. În asemenea cazuri, apariția dovezii este explicată cel mai bine de ipoteză, dar faptul că dovada există este folosit pentru a veni chiar în sprijinul ipotezei”.

54. Mai este aici o particularitate tehnică, dezvăluind relația profundă dintre autoorganizare, teoria informației și inferența existențială. Surprinderea nu este decât verosimilitate cu logaritm negativ a unor stări senzoriale, dat fiind un model în cadrul unei zone Markov. În statistica bayesiană, acest lucru e cunoscut și sub numele (logaritm de) verosimilitate marginală (*model evidence*). Asta permite afirmaţia că minimizarea energiei libere Friston *se autovalidează*, adică maximizarea dovezilor pentru un model care generează stările senzoriale ale oricărui sistem ce are o zonă Markov.

55. De aceea, ar putea fi mai corect să spunem: „Răspunsul la această întrebare trebuie să determine comportamentul mediu al sistemului”.

56. Observație: acest mecanism impune constrângeri

probabiliste „liberului-arbitru”. De exemplu, ești liber să intri în cușca leului, dar e *improbabil* că vei intra și, dacă intri, *probabil* vei muri. Vezi, mai jos, teorema lui Bayes.

57. Unde *A* și *B* sunt evenimente, *P*(*A* | *B*) este verosimilitatea ca *A* să apară dat fiind că *B* este adevărat (probabilitatea condițională), *P*(*B* | *A*) este verosimilitatea ca *B* să apară dat fiind că *A* este adevărat (altă probabilitate condițională), iar *P*(*A*) și *P*(*B*) sunt probabilitățile de a observa *A* și *B* independent unul de celălalt (probabilitatea marginală). Această formulare a teoremei e conformă cu Joyce (2008), care a scris *P*(*A*) | *P*(*B*) ca raport separat.

58. Teorema lui Bayes poate fi reformulată în termenii energiei libere – prin care energia liberă este descompusă în „acuratețe” și „complexitate”. Verosimilitatea marginală este diferența dintre acuratețe și complexitate, întrucât modelele cu energie liberă minimă furnizează explicații precise ale datelor în contextul unor costuri ale complexității, ceea ce, apoi, înseamnă că reducerea complexității modelului îmbunătățește „generalizabilitatea” modelului, dar cu costul acurateței. (În termeni bayesieni, „verosimilitatea” trebuie evaluată în raport cu „probabilitatea” pentru a preveni supraajustarea.) Knill și Pouget (2004), p. 713, ajung la esența problemei: „Adevăratul test al ipotezei codificării bayesiene constă în a vedea dacă estimările neuronale ce duc la evaluări perceptive sau la comportament motor țin seama de incertitudinea existentă în fiecare etapă a prelucrării”.

59. Pentru o descriere detaliată a microanatomiei funcționale a acestui proces, vezi Friston (2005), Adams, Shipp și Friston (2013) și Parr și Friston (2018).

60. Astfel, de exemplu, erorile de predicție par să fie

comunicate folosind frecvențe (înalte) din banda gama, iar predicțiile par să fie transmise prin frecvențe (joase) din banda beta. Vezi Bastos *et al.* (2012, 2015).

61. Vezi Tozzi, Zare și Benasich (2016): „Minimizarea energiei libere variaționale presupune cu necesitate o codificare eficientă sub aspect metabolic corespunzătoare principiilor redundanței minime și transferului informațional maxim (Picard și Friston, 2014). Maximizarea informației reciproce și minimizarea costurilor metabolice sunt două fețe ale aceleiași medalii: prin descompunerea energiei libere variaționale în acuratețe și complexitate, putem deduce principiul informației reciproce maxime ca un caz special al maximizării acurateței, iar minimizarea complexității se traduce prin minimizarea costurilor metabolice (Friston *et al.*, 2015). Astfel, forma de bază a principiului energiei libere al lui Friston susține ideea că nivelurile energetice ale activității cerebrale spontane, care sunt mai mici în comparație cu activitatea evocată, permit SNC să aibă două realizări aparent contradictorii: să minimizeze cât mai mult posibil costurile metabolice și să maximizeze cât de mult se poate informația reciprocă”.

62. Vezi Clark (2015), p. 268. Totuși, acest lucru trebuie comparat cu ceea ce spune el despre activitatea implicită a sistemului CĂUTĂRII (p. 263): „Asemenea creaturi sunt construite să-și caute parteneri, să evite foamea și setea și să se implice (chiar dacă nu le este foame sau sete) în tipurile de explorare sporadică a mediului înconjurător care le ajută să se pregătească pentru schimbări ambientale neprevăzute, penurii de resurse, concurenți noi și așa mai departe. Așadar, în orice moment, eroarea de predicție este

minimizată numai în contextul acestui set complex de «preconizări» ce definesc creatura”.

63. Friston și Stephan (2007), p. 427.

64. M-am inspirat pentru această fabulă din metafora lui Clark (2015) despre un baraj cu fisuri.

Capitolul 8. O ierarhie predictivă

1. Hohwy (2013), p. 63.

2. Spun „centrul”, și nu „vârful” ierarhiei deoarece pare ciudat (din punct de vedere anatomic) să plasezi nucleii hipotalamici și ai trunchiului cerebral care monitorizează corpul *deasupra* neocortexului. Prefer să vizualizez ierarhia ca desfășurându-se concentric, dinspre interior spre exterior, cumva ca foile unei cepe (vezi Mesulam, 2000). Până la urmă, după cum știe orice embriolog, sistemul nervos este un tub. Să nu uităm nici că tubul neural se formează din ectoderm, prin invaginarea plăcii neurale, prin care canalul central ia locul lumii exterioare.

3. Aceasta nu este perspectiva standard asupra codificării predictive, dar tocmai de asta am și scris cartea de față: arăt cum anume perspectiva noastră asupra creierului este îngreunată de cortico-centrism. *Cf.* Pezzulo (2014), p. 910: „O direcție interesantă pentru cercetările viitoare ar fi analizarea reglării homeostatice în cadrul inferenței active”.

4. Clark (2015), p. 21, oferă un scurt rezumat, folosind vederea ca model: „Un anumit tipar de stimulare retiniană, întâlnit într-un context dat, ar putea fi descris cel mai bine folosind un model generativ care [...] combină reprezentări de nivel

superior ale unor agenți, obiecte, motive și mecanisme care interacționează, cu multe straturi intermediare ce surprind modul în care culorile, formele, texturile și marginile se combină și evoluează în timp. Când combinația unor asemenea cauze ascunse (ce se întind pe multe scări spațiale și temporale) se stabilizează într-un tot coerent, sistemul autogenerează datele senzoriale folosind cunoașterea stocată și percepe o scenă structurată ce are semnificație. Merită să spunem din nou că această perspectivă asupra scenei distale structurate trebuie să fie generată folosindu-se numai informațiile disponibile din perspectiva animalului. Adică ea trebuie să fie o perspectivă înrădăcinată în totalitate în combinarea dintre orice prestructurare (a creierului și corpului) care ar putea fi prezentă datorită istoriei evoluției animalului și influențele stimulării energetice consemnate de receptorii senzoriali. Un mijloc sistematic de a obține asemenea perspectivă este furnizat de încercarea continuă de a autogenera semnalul senzorial folosind o arhitectură multinivel. Concret, asta înseamnă că unele conexiuni de sus în jos și laterale dintr-un sistem multinivel ajung să codifice un model probabilistic de cauze ce interacționează și funcționează la mai multe scări în spațiu și timp. Recunoaștem obiecte, stări și situații [...] prin identificarea setului de factori ce interacționează (cauze distale) în cazul cărora există cea mai mare probabilitate ca combinarea lor să genereze (de aici, prezice și explică cel mai bine) datele senzoriale de intrare”.

5. *Cf.* Gregory (1980).
6. Acest paragraf îl parafrazează pe Hohwy (2013), pp. 81-82, dar îi corectează biasul exteroceptiv.

7. Nu vreau să spun că punctele de stabilizare homeostatică nu se schimbă *niciodată*, ci doar că gradele lor de libertate sunt extrem de limitate; prin urmare, aproape întreaga alostază implică schimbare comportamentală, și nu actualizarea elementelor anterioare înnăscute.
8. Trec aici peste chestiunea unui „Bayes empiric”. E posibil să învățăm de la zero, dar e îngrozitor de costisitor din punct de vedere biologic.
9. Ca să ilustrez dimensiunile problemei, să reflectăm la cuvintele lui Hohwy (2013), p. 206: „Deși ne spune întreaga poveste a erorii de predicție, nu pare să existe nicio contradicție în ideea că există o creatură dotată cu tot aparatul necesar pentru minimizarea erorii de predicție, care se implică în aceeași inferență predictivă ca și noi – indiferent de gradul de detaliere naturală, neuronală pe care vrem să-l specificăm –, dar care nu este conștientă din punct de vedere fenomenologic. Desigur, ne-am aștepta ca această creatură să fie conștientă, dar nimic din toată povestea ei fizică nu impune să și fie conștientă. Asta înseamnă că rămâne deschisă întrebarea dacă conștiința este ceva dincolo de ce este fizic sau nu”.

   Cum ar putea să *nu* fie conștientă o creatură având exact același aparat neuronal ca tine și ca mine? Îndrăznesc să afirm că „zombiului filosofic” imaginat aici de Hohwy îi lipsește aparatul pentru sentimente. Pentru o interpretare alternativă a problemei pe care o pune el, vezi capitolul 12.

Capitolul 9. De ce și cum se naște conștiința

1. Rolls (2019), p. 10, sublinierea mea. Vezi și comentariul meu: Solms (2019b).
2. „Nu trebuie folosite mai multe lucruri decât e necesar.” Dacă există mai multe moduri posibile de a explica ceva, cel care folosește cele mai puține presupuneri este probabil cel corect.
3. Roger Sessions, *New York Times*, 8 ianuarie 1950.
4. Friston (2009), p. 299, definește rolul preciziei ca fiind de „a controla influența relativă a așteptărilor anterioare la diferite niveluri”. După cum arată Hohwy (2013), p. 199: „Dacă se crește amplificarea [precizia] unui semnal, atunci trebuie redusă amplificarea altor semnale. Altfel, ideea de amplificare își pierde semnificația: suma ponderilor trebuie să fie unu. Așadar, pe măsură ce așteptările pentru precizie cresc amplificarea unei erori de predicție, amplificarea altora se va reduce”. *Cf.* Clark (2016), p. 313: „Feldman și Friston (2010) arată că precizia se comportă ca și cum ar fi ea însăși o resursă limitată, în sensul că creșterea preciziei pentru unele erori de predicție necesită reducerea ei pentru altele. De asemenea, e interesant comentariul autorilor (*op. cit*., p. 11): «Motivul pentru care precizia se comportă ca o resursă este că modelul generativ conține credințe anterioare conform cărora precizia logaritmică este redistribuită prin canale senzoriale într-o modalitate sensibilă la context, dar este conservată în toate canalele»”. Aceste afirmații amintesc de cele speculate de mine în capitolul 6 despre fiecare neuromodulator ce ajustează diferite aspecte ale activării. E interesant că oscilațiile (erorile) gama răspund la acetilcolină.
5. Neurofiziologii numesc precizia și „amplificare”.

Terminologia foarte variată folosită de specialiști din diverse discipline poate fi derutantă aici. „Activitatea” sinaptică (= neurotransmisia) este modulată de „amplificarea” postsinaptică (= neuromodularea), care, în timp, determină „eficiența” (= neuroplasticitatea). Aceleași variabile sunt descrise de specialiștii în științe computaționale drept „stări” de semnal, „precizii” de semnal și, respectiv, „parametri” de semnal. În general, neurotransmisia necesită stări, neuromodularea necesită precizii și plasticitatea necesită parametri.

6. Friston numește acest lucru „căutare epistemică a hranei”.

7. Acest paragraf îl parafrazează îndeaproape pe Clark (2015), p. 70.

8. De aici apare „orbirea perceptivă”, cel mai cunoscut exemplu găsindu-se pe https://www.youtube.com/watch?v=vJG698U2Mvo (vezi Chabris și Simons, 2010).

9. Varela, Thompson și Rosch (1991), p. 198. Varela numește „institutivă” această abordare a percepției. Vezi Clark (2015), p. 173: „Preocuparea generală a unei abordări institutive a percepției nu este să determine cum urmează să fie recuperată o lume independentă de observator, ci să determine principiile comune sau legăturile legitime dintre sistemele senzoriale și motoare ce explică modul în care acțiunea poate fi ghidată perceptiv într-o lume dependentă de observator”.

10. Există un volum mare de literatură experimentală care susține empiric această conceptualizare formală a preciziei. Deocamdată voi menționa un singur studiu, care ilustrează ideea de bază. Folosind RMNf, Hesselmann și colaboratorii săi (2010) au analizat activitatea de stare inițială din creier

în condiții care manipulau așteptările legate de precizie. Autorii le-au prezentat participanților la cercetare stimuli (vizuali și auditivi) cu diferite niveluri de zgomot; când participanții se așteptau la precizie, a fost crescută amplitudinea, iar când se așteptau la imprecizie, amplitudinea a fost redusă. Adică predicția de sus în jos a fost facilitată atunci când erau preconizate semnale senzoriale imprecise. Vezi Feldman și Friston (2010) pentru o descriere completă a modului în care optimizarea preciziei explică atât atenția endogenă, cât și atenția exogenă.

11. După cum îți amintești de la p. 45, Freud a scris următoarele la vremea respectivă, într-o scrisoare către Fliess, din 20 octombrie 1895: „Într-o noapte plină [...] barierele au fost brusc ridicate, vălurile au căzut și am putut să înțeleg de la detaliile nevrozelor la determinanții conștiinței. Totul părea să se potrivească, rotițele erau angrenate, obiectul îți dădea impresia că era într-adevăr o mașină care, curând, va merge de la sine”. Și a continuat: „Cele trei sisteme de neuroni [ϕ, ψ și ω], condițiile de cantitate libere și fixe [$Q\eta$], procesele primare și secundare, tendința principală și tendința de compromis a sistemului nervos, cele două reguli biologice ale atenției și apărării, indicatorii calității [ω], realității și gândirii, starea grupurilor psihosexuale, determinarea sexuală a represiei și, în final, determinanții conștiinței ca funcție perceptivă – toate acestea se potriveau și încă se potrivesc! Desigur, nu îmi pot ascunde încântarea”. Totuși, la scurt timp, a realizat că rotițele nu erau angrenate, lucrurile nu se potriveau complet. Atunci a abandonat proiectul. Când eu și Friston l-am reînviat – și, luând în calcul mai bine de un secol de progrese în neuroștiințe, sperăm că l-am

finalizat (cel puțin în mare; vezi Solms, 2020b) –, mi-am amintit de remarca dură a lui Freud: „Obiectul îți dădea impresia că era într-adevăr o mașină care, curând, va merge de la sine”. Vezi capitolul 12.

12. De observat că ω nu este un vector, ci un scalar (măsoară o matrice a preciziei). De observat și că termenii Q and $Q\eta$ au fost utilizați ușor diferit în Solms și Friston (2018) față de cum îi folosesc eu aici.
13. Unde e este un vector, iar F un scalar. Notația cu puncte în ecuațiile de mai jos implică un produs scalar (adică înmulțire a matricii sau a vectorului). Aceste ecuații reprezintă în întregime contribuția lui Friston la colaborarea noastră.
14. $F \approx -\log P(\phi(M))$.
15. $E[F] \approx E[-\log P(\phi)] = H[P(\phi)] = -½ \cdot \log(|\omega|)$, unde $E[\cdot]$ denotă preconizarea sau media, iar P și H probabilitatea și, respectiv, entropia (ca și înainte); în baza premiselor gaussiene despre fluctuația aleatorie.
16. Importanța generală a optimizării preciziei a fost recunoscută cu mult timp în urmă, mai ales în legătură cu atenția (vezi Friston, 2009; Feldman și Friston, 2010). Totuși, primul specialist care i-a recunoscut importanța pentru conștiință a fost Katerina Fotopoulou (2013), p. 35: „Un aspect esențial al conștiinței poate servi la consemnarea proprietății anterior menționate a «incertitudinii» și a proprietății sale inverse, precizia. Această opinie contrazice vechea perspectivă intuitivă asupra conștiinței afective esențiale ca proprietate hedonică de monitorizare, exprimată de Solms în termeni freudieni ca seria plăcere-neplăcere. În schimb, eu afirm că proprietatea esențială a

acestui aspect al conștiinței (ca fiind opus conștiinței perceptive [...]) este un fel de principiu al certitudinii-incertitudinii sau al dezambiguizării". Țin să precizez că, în ce privește plăcerea-neplăcerea versus certitudinea-incertitudinea, cred că Fotopoulou m-a înțeles puțin greșit, de unde și răspunsul meu: „Nu sunt sigur că [...] înțeleg la ce se referă când afirmă că afectul monitorizează incertitudinea, și nu proprietatea hedonică. În opinia mea, proprietatea hedonică *este* măsura incertitudinii" (Solms, 2013, p. 81).

17.

$$\frac{\partial}{\partial t}M = -\frac{\partial F}{\partial M} = -\frac{\partial F}{\partial e}\frac{\partial e}{\partial M} = \frac{\partial \varphi}{\partial M}\cdot\omega\cdot e \quad (1)$$

$$\frac{\partial}{\partial t}Q\eta = -\frac{\partial F}{\partial Q\eta} = -\frac{\partial F}{\partial e}\frac{\partial e}{\partial Q\eta} = \frac{\partial \psi}{\partial Q\eta}\cdot\omega\cdot e \quad (2)$$

$$\frac{\partial}{\partial t}\omega = -\frac{\partial F}{\partial \omega} = \frac{1}{2}\cdot(\omega^{-1} - e\cdot e) \quad (3)$$

Unde ∂ denotă un derivat parțial și *t* timpul, iar eroarea de predicție și energia liberă sunt:

$$e = \varphi(M) - \psi(Q\eta)$$

$$F = \frac{1}{2}\cdot(e\cdot\omega\cdot e - \log(\omega))$$

Prezint în mare aceste ecuații, ca niște desene cu pretenții. Ele trebuie să fie dezvoltate suplimentar, iar această elaborare se poate face numai atunci când sunt implementate (vezi capitolul 12). De exemplu, după cum am afirmat mai sus, în abordări mai complexe, ar trebui să avem în vedere și modelele generative ierarhice (cu precizii la fiecare nivel) și să includem incertitudinea condițională despre stările externe. Mai mult, ecuațiile pun laolaltă toate erorile senzoriale de predicție – incluzând modalitățile exteroceptive, proprioceptive și interoceptive. De observat că, aici, termenul „proprioceptiv" este sinonim cu

„chinestezic” (Friston folosește „proprioceptiv” pur și simplu pentru armonia aliterativă cu „exteroceptiv” și „interoceptiv”).

18. În termeni tehnici, acest lucru se numește „algoritmul gradientului descendent”, unde gradientul este rata de schimbare a energiei libere cu precizie.
19. În baza asumpțiilor noastre simplificatoare despre codificarea credințelor bayesiene (vezi mai sus).
20. În acest sens, se poate spune că precizia joacă rolul „demonului lui Maxwell” – un experiment mental creat de fizicianul James Clerk Maxwell (1872): un demon controlează o ușă mică între două încăperi cu gaze. Moleculele de gaz plutesc cu viteze diferite. Când cele mai rapide din prima încăpere ajung la ușă, demonul o deschide și o închide, pentru foarte scurt timp, astfel încât ele trec în a doua încăpere, iar cele mai lente rămân în urmă. Întrucât moleculele mai rapide generează mai multă căldură decât cele mai lente, acest lucru va scădea entropia – ceva ce nu se poate întâmpla fără lucru mecanic. Dacă punem semnul egalității între trecerea moleculelor din această analogie și neurotransmisia semnalelor senzoriale, atunci cântărirea preciziei (neuromodularea) face ce face demonul lui Maxwell: selectează semnale senzoriale pentru a încurca a doua lege a termodinamicii. De observat că, în termenii dinamicii sistemului descrise aici, conștiința este *activitatea demonului lui Maxwell*, și nu *trecerea moleculelor* facilitată de activitate. Asta înseamnă că conștiința este *optimizarea preciziei în raport cu energia liberă*, și nu *trecerea mesajelor printr-o ierarhie predictivă*. Prin urmare, în figura 17 conștiința este activitatea ω (precizia), care determină

influența relativă a *e* (semnalele de eroare) asupra *Qη* (modelul intern). Demonul-precizie modelează influența erorilor în legătură cu modelul. Conștiința depinde de un atribut al credințelor, și nu de conținutul credințelor (adică depinde de precizia fluctuantă a credințelor despre situațiile interne și externe [sau de încrederea în ele]). Activitatea acestui demon *cauzează* sechele senzoriale prin amplificarea sau atenuarea erorilor de predicție; optimizarea preciziei nu este inerentă semnalelor de eroare în sine. Pentru mai multe despre implicațiile biologice ale demonului lui Maxwell, vezi excelenta carte a lui Paul Davies (2019) care, din păcate, a apărut după ce am realizat noi aplicarea conceptului la conștiință (Solms și Friston, 2018), așa că nu am citat-o.

21. Brown *et al.* (2013), Feldman și Friston (2010), Frith, Blakemore și Wolpert (2000).

22. Cisek și Kalaska (2010), Frank (2005), Friston *et al.* (2012), Friston, Schwartenbeck, FitzGerald *et al.* (2014), Moustafa, Sherman și Frank (2008).

23. Hohwy (2013), Seth (2013), Ainley *et al.* (2016). Pentru sensibilitatea interoceptivă și modularea socială a durerii, vezi și Crucianelli *et al.* (2017), Fotopoulou și Tsakiris (2017), Krahé *et al.* (2013), Decety și Fotopoulou (2015), Paloyelis *et al.* (2016) și Von Mohr și Fotopoulou (2017).

24. Ferrarelli și Tononi (2011), Lisman și Buzsaki (2008), Uhlhaas și Singer (2010).

25. Vezi Hobson (2009) și Hobson și Friston (2012, 2014). Totuși, în opinia mea, o explicație satisfăcătoare a viselor în acest cadru trebuie să înceapă de la caracterul lor *conștient* (și afectiv).

26. Nour și Carhart-Harris (2017).
27. Dehaene și Changeux (2011), Friston, Breakspear și Deco (2012).
28. Montague *et al.* (2012), Corlett și Fletcher (2014), Friston, Stephan, Montague și Dolan (2014), Wang și Krystal (2014).
29. Pentru o discuție despre „empatie”, vezi Solms (în tipar). Termenul german este *Einfühlung*, care literal înseamnă „simțire în”.

Capitolul 10. Înapoi la cortex

1. Zeki (1993), p. 236, sublinierea mea.
2. *Ibidem*, p. 238.
3. Nu intenționez să implic că acesta este un proces cognitiv de nivel superior. Cortexul vizual „de proiecție” pare să umple pata oarbă cu orice o înconjoară. În condiții naturale, la asta contribuie mișcările frecvente ale ochilor, care se asigură că ai aproape mereu o imagine reală (memorie pe termen ultrascurt) a ceea ce este în pata oarbă. Pentru diversitatea mecanismelor implicate, vezi Ramachandran (1992), Ramachandran și Gregory (1991) și Ramachandran, Gregory și Aiken (1993). Se întâmplă că, pe lângă umplerea cu conținut a petei oarbe, are loc și eliminarea obiectelor nedorite, cum ar fi umbrele „musculițelor” entoptice și vascularizația retiniană.
4. Suprapunerea dintre câmpurile reprezentate de fiecare ochi anulează petele oarbe, ceea ce înseamnă că trebuie să închizi un ochi ca să observi umplerea iluzorie menționată.
5. Există, desigur, o anumită suprapunere între ele, dar măsura

în care obiectele atenției vizuale sunt literalmente secționate în două este ușor de demonstrat în cazurile de neglijare unilaterală (stânga) în urma unor leziuni la nivelul emisferei drepte.

6. Întâmplător, aceasta este o dovadă pentru opinia că funcția „de legare” a conștiinței vine de sub cortex, din trunchiul cerebral unitar, și nu din cele două emisfere cerebrale (vezi Panksepp și Biven, 2012).

7. Solms *et al.* (1988).

8. L-a descris drept „un teatru minunat” (Helmholtz, 1867, p. 776).

9. *Ibidem*, p. 438.

10. Majoritatea oamenilor care lucrează în paradigma codificării predictive nu acordă suficientă atenție faptului că inferența perceptivă este un proces *inconștient*. De exemplu, eu nu sunt de acord cu Hohwy (2013) când spune că „suntem conștienți de «fantezia» generată de modul în care predicțiile curente atenuează eroarea de predicție”. În opinia mea, nu *devenim conștienți* de „fanteziile” predictive dacă nu *se ciocnesc* cu realitatea. Și slavă Domnului pentru acest lucru.

Se pare că Clark este de acord cu mine în această privință (Lupyan și Clark, 2015, p. 281, sublinierea mea): „Deși majoritatea predicțiilor sunt inconștiente, uneori le putem conștientiza *când sunt încălcate.* De exemplu, imaginează-ți că bei dintr-un pahar în care crezi că este suc de portocale, dar realizezi, după ce îl guști, că, de fapt, e lapte. Diferența dintre gustul de lapte când te aștepți să fie lapte și când, în loc de lapte, te aștepți să fie suc de portocale este conștientizarea așteptării referitoare la sucul de portocale

(Lupyan, 2015). În mod similar, gândește-te la experiența unei omisiuni nepreconizate, ca atunci când lipsește o notă muzicală dintr-o compoziție familiară. Asemenea omisiuni pot fi la fel de frapante sub aspect perceptiv și la fel de predominante ca tonul cel mai vibrant – altfel, un efect uimitor ce se explică clar dacă presupunem că construcția experienței perceptive implică așteptări bazate pe vreun fel de model a ceea ce e probabil să se întâmple". În pofida afirmațiilor lui Hohwy (2013), uneori el se apropie de opinia mea. Într-un articol anterior, a făcut următoarea afirmație (Hohwy, 2012, p. 11, sublinierea mea): „Această semnătură temporală corespunde codificării predictive în măsura în care, când eroarea de predicție a unui stimul este suprimată complet și nu mai are loc nicio explorare (întrucât inferența activă este atenuată, din cauză că sistemul e ocupat temporar pe parcursul atenției defocalizate), probabilitatea ar trebui să înceapă să scadă. Ajungem aici pornind de la ideea că *ceea ce impulsionează percepția conștientă este procesul concret al suprimării erorii de predicție*". Totuși, în cartea lui (2013, p. 201, sublinierea mea), spune: „Percepția conștientă este efectul inferenței perceptive inconștiente. Nu ne implicăm în mod conștient într-o actualizare bayesiană a predicțiilor noastre anterioare în lumina unor dovezi noi și nici a modului în care inputul senzorial este prezis și apoi atenuat. Ceea ce este conștient e rezultatul inferenței – *concluzia*". De aceea nu sunt de acord cu Hohwy (2013), dacă înțeleg bine ce vrea să spună. Dar suntem amândoi de acord că *travaliul* constând în minimizarea semnalelor de eroare precise generează conștiința, adică discrepanța problematică dintre predicție și eroarea cântărită de

precizie. Dar nu sunt de acord cu el când spune că „suntem conștienți de «fantezie»”. În cel mai bun caz, Hohwy pare să creadă că suntem conștienți de „fantezia” ce încearcă *să suprime* eroarea de intrare, dar eu cred că suntem conștienți de faptul că *nu* suprimă eroarea – adică suntem conștienți de „lucrul în desfășurare” predictiv cauzat de discrepanță. Așa devine realitatea predominantă. Poate că această idee este doar o chestiune de semantică. Pentru mine, contează că conștiința în percepție este impulsionată de *incertitudine*, și nu de *cele-mai-bune-presupuneri* generate de certitudinea relativă. După cum văd eu situația, presupunerile devin conștiente numai atunci când sunt nesigure și se retrag când sunt confirmate. Cu alte cuvinte, devenim conștienți de fanteziile noastre numai atunci când sunt contrazise de realitate. Conștiința ar putea fi descrisă ca un proces de *dezambiguizare*. Sunt conștient că această chestiune este complicată de diferența dintre inferența perceptivă și cea activă (și dintre atenția exogenă și cea endogenă). Așa se explică de ce există o literatură bogată – cea pe care se bazează Hohwy (2013) – care sugerează că percepem în mod preferențial (conștient) ceea ce se potrivește cu așteptările noastre *și* există și alte lucrări de specialitate care sugerează că percepem în mod preferențial (conștient) ceea ce este cel mai nepreconizat. Pentru mine, această contradicție este rezolvată de conceptul *prioritizării afectului*, care este impulsionată, în ambele cazuri, de echilibrul dintre ceea ce eu numesc în sens larg „nevoi” și „oportunități” (vezi capitolul 5). Pe scurt, percepem conștient ceea ce este predominant *în relație cu nevoia prioritizată pe moment*. Nevoile (afectele) prioritizate produc, prin definiție, cele mai

precise semnale de eroare. Opiniile mele în această privință sunt elaborate în secțiunea următoare. Totuși, în cele din urmă, aceasta este o problemă empirică. În prezent, împreună cu studenții mei Donne van der Westhuizen și Julianne Blignaut și cu fostul meu student Joshua Martin, o studiez folosind paradigma rivalității binoculare standard, ce se referă la *conștiința* perceptivă. Vezi și Pezzulo (2014), Yang, Zald și Blake (2007) și Stein și Sterzer (2012).

11. Am publicat pentru prima dată această idee în Solms (2013), apoi am revenit asupra ei în diverse moduri, de exemplu, Solms (2015b, 2017b, 2017c, 2018b).
12. Freud (1920), p. 25.
13. Freud a atribuit conștiință și memorie unor sisteme diferite de neuroni (ω și, respectiv, ψ) care, ulterior, au devenit sistemele lui metapsihologice Cs și Pcs – dar a considerat că *amândouă* sunt sisteme corticale. Pasajul din care este citată fraza „conștiința ia naștere acolo unde se sfârșește urma mnezică” afirmă foarte clar acest lucru, deși a fost scris la 25 de ani după „Proiect” (vezi Freud, 1920, pp. 25 și urm.).
14. Bargh și Chartrand (1999), p. 476: „Unele dintre sistemele de ghidare automată pe care le-am evidențiat sunt «naturale» și nu necesită experiență pentru a se dezvolta. Acestea sunt, pe de o parte, asocierea reprezentărilor perceptive și comportamentale și conexiunea dintre procesele de evaluare automată și, pe de altă parte, dispoziția și comportamentul. Alte forme de autoreglare automată apar din experiența repetată și consecventă. Ele se conectează la regularitățile experienței individului și preiau sarcini de la ghidarea și alegerea conștiente atunci când alegerea respectivă nu este exercitată cu adevărat. Astfel, obiectivele

și motivele pot ajunge să funcționeze nonconștient în anumite situații, stereotipurile pot fi asociate multă vreme cu caracteristicile perceptive ale unor grupuri sociale, iar evaluările pot fi integrate cu reprezentările perceptive ale persoanei, obiectului sau evenimentului astfel încât devin active imediat și neintenționat în cursul percepției”.

15. Din punct de vedere tehnic, stările neuronale, care sunt intensificate de preciziile neuronale, actualizează parametrii neuronali.

16. Asta coincide perfect cu ideea lui Freud de „proces secundar”. El a conceput procesele secundare drept o „legare” a formei primare de energie pulsională, care este „mobilă”. Legarea energiei libere Friston reprezintă fundamentul mecanic a ceea ce numim *travaliu* psihic (lucru mecanic util). Această concluzie este foarte importantă: „Eu cred că această deosebire [dintre energia liberă și energia latentă] prezintă până acum înțelegerea cea mai profundă pe care o avem despre natura energiei nervoase și nu văd cum s-ar putea ca cineva s-o evite” (Freud, 1915b, p. 188) [ed. rom.: „Inconștientul”, traducere de G. Lepădatu, în *Opere esențiale*, vol. 3: *Psihologia inconștientului*. București: Trei, e-book]. De reținut totuși că Freud considera că procesul secundar poate să funcționeze și *în mod preconștient*. De aici apare o întrebare empirică: poate procesarea corticală să-și îndeplinească funcția de stabilizare în lipsa conștiinței? Această întrebare amintește de controversa actuală cu privire la memoria de lucru nondeclarativă (vezi Hassin *et al.*, 2009).

17. Nu vreau să sugerez că orice reconsolidare aduce cu sine conștientizarea *memoriei aflate în curs de actualizare* – ea

poate aduce cu sine modularea afectivă a procesului de actualizare ce rămâne inconștient, din punct de vedere cognitiv (vezi mai jos). Se întâmplă că mecanismele pe care tocmai le-am menționat explică fenomenul psihologic pe care Freud l-a numit „rezistență”: reticența noastră aparte de a ne actualiza modelele predictive în fața unor dovezi contradictorii. (Din păcate, această afirmație e valabilă chiar și în cazul unor specialiști!)

18. Termenul „recompensă = predicție” din diagrama lui Schultz i-ar putea deruta pe unii cititori. Atunci când consecințele senzoriale ale unei acțiuni se potrivesc cu consecințele prezise, nu se întâmplă nimic, nu există nicio „recompensă”. Senzațiile, atât negative, cât și pozitive, semnifică întotdeauna erori (vezi figura 12). Totuși, în limbaj behaviorist, „recompensă” nu implică vreo senzație. Înseamnă doar că predicția este consolidată. Adică, folosind terminologia mea, îi este atribuită o precizie mai mare. Pentru mine, asta înseamnă că senzația va fi resimțită drept plăcută *dacă predicția relevantă a fost prioritizată de triunghiul de decizie din trunchiul cerebral.*
19. Misanin, Miller și Lewis (1968).
20. Nader, Schafe și LeDoux (2000). Vezi Dudai (2000) pentru o prezentare accesibilă. Reconsolidarea este strâns legată de conceptul de „retranscriere” a memoriei în accepțiunea lui Freud. Vezi scrisoarea lui către Fliess datată 6 decembrie 1896, în care anticipează, cel puțin, conceptul de consolidare a sistemelor: „După cum știi, lucrez la ipoteza că mecanismul nostru fizic ia naștere printr-un proces de stratificare: materialul prezent sub forma unor urme mnezice care este supus din când în când unei *rearanjări*

corespunzătoare unor condiții noi – supus unei *retranscrieri*. Astfel, nouă la teoria mea este teza că memoria e prezentă nu o dată, ci de mai multe ori, că este exprimată în diferite tipuri de indicații. Am postulat un tip similar de rearanjare cu ceva timp în urmă (*Afazia* [Freud, 1891]) pentru căile ce pleacă din periferie. Nu pot să spun câte asemenea consemnări există: cel puțin trei, probabil mai multe. Asta se vede în următoarea imagine schematică [în care conștiința apare drept «consemnarea» finală a urmei; vezi figura 8)], în care se presupune că consemnările diferite sunt și separate (nu neapărat topografic) în funcție de neuronii ce le sunt vehicule. Poate că această ipoteză nu e necesară, dar e cea mai simplă și poate fi acceptată provizoriu". Freud a conceptualizat „represia" ca pe un eșec de retranscriere. Vezi Solms (2017c), pentru o actualizare din perspectivă neuropsihanalitică.

21. Astfel, atât potențarea pe termen lung, cât și depresia pe termen lung sunt modulate de sistemul reticulat activator (vezi Bienenstock, Cooper și Munro, 1982). De unde și capacitatea ECT și a crizelor epileptice, toate acționând prin intermediul sistemului reticulat activator, de a interfera cu consolidarea memoriei.

22. De fapt, Hebb (1949), p. 62, a spus: „Să presupunem că persistența sau repetarea unei activități cu reverberație (sau «urmă») tinde să inducă modificări celulare de durată care contribuie la stabilitatea ei [...]. Atunci când un axon al celulei *A* este suficient de aproape încât să excite o celulă *B* și ia parte în mod repetat sau persistent la descărcarea ei, are loc un anumit proces de creștere sau o schimbare metabolică într-o celulă sau în amândouă, astfel încât

eficiența celulei *A*, ca una dintre celulele care o descarcă pe *B*, crește”.

23. În prezent, planific împreună cu Cristina Alberini o serie de experimente privind rolurile PAG și ale sistemului reticulat activator în învățarea conștientă versus învățarea inconștientă. Aceste experimente vor clarifica rolurile diverșilor nuclei „ai activării” din trunchiul cerebral superior care modulează cogniția conștientă din prozencefal. Paradigma reconsolidării promite să dezvăluie unele dintre mecanismele intracelulare elementare ale conștiinței perceptive în relație cu învățarea. De exemplu, atât consolidarea, cât și reconsolidarea pot fi perturbate de inhibarea sintezei proteice și amândouă necesită factorul CREB de transcripție a genelor. Totuși, cercetările recente sugerează că în nucleul amigdalian este necesar factorul neurotrofic derivat din creier (BDNF) pentru consolidare, dar nu și pentru reconsolidare, precum și că Zif268, factorul de transcripție și genă precoce imediată, este necesar pentru reconsolidare, dar nu și pentru consolidare. O dublă disociere similară între Zif268 pentru reconsolidare și BDNF pentru consolidare s-a descoperit în hipocamp. Vezi Debiec *et al.* (2006) și Lee, Everitt și Thomas (2004).

24. Riggs și Ratliff (1951), Ditchburn și Ginsborg (1952). E instructiv să ne amintim observațiile lui Helmholtz despre atenție (1867, p. 770): „Starea naturală neforțată a atenției noastre este de a hoinări continuu spre alte lucruri, astfel încât, atunci când se epuizează interesul față de un obiect, când nu mai putem percepe nimic nou, atunci, împotriva voinței noastre, atenția ni se îndreaptă spre altceva [...]. Dacă vrem să ne menținem atenția asupra unui obiect,

trebuie să tot găsim ceva nou la el, mai ales dacă alte senzații puternice încearcă să o deturneze”. Aceste afirmații trebuie puse în legătură cu am spus mai sus despre CĂUTAREA implicită.

25. Vezi https://en.wikipedia.org/wiki/Lilac_chaser

26. Dacă îți spune cineva că aceste efecte nu implică creierul (cu alte cuvinte, că ele apar numai la nivelurile periferice ale sistemului nervos), zi-i să citească Hsieh și Tse (2006). Vezi și Coren și Porac (1974).

27. Vezi Oberauer *et al.* (2013).

28. Exclud de aici „amorsajul” și învățarea perceptivă. Acestea necesită într-adevăr imagini și implică cortexul. Majoritatea celorlalte afirmații pe care le fac despre memoria nondeclarativă nu sunt valabile în cazul amorsajului și învățării perceptive, care aparțin altei categorii; ele sunt „eșafodajul” memoriei declarative. Vezi ce spun mai departe despre funcția de amorsaj a cuvintelor, de exemplu.

29. Ganglionii bazali nu sunt afectați la unii copiii cu hidroanencefalie (dar nu la toți) și la unele animale supuse procedurii de decorticare cerebrală.

30. Din punct de vedere tehnic, complexitatea este entropia relativă dintre credințele posterioare și cele anterioare sau dintre distribuțiile probabilității în stările externe. Această definiție a complexității decurge din faptul că verosimilitatea marginală este diferența dintre acuratețe și complexitate (vezi capitolul 8). Întrucât verosimilitatea marginală este sporită în mod activ prin minimizarea energiei libere, acuratețea predicțiilor crește, ca și complexitatea lor. Cu alte cuvinte, creșterea complexității modelului este întotdeauna permisă de capacitatea de a face

predicții mai precise – cum se întâmplă de regulă în sistemele de memorie corticale.

31. Compară cu Hohwy (2013), p. 202: „Ideea ar fi că inferența perceptivă circulă într-un spațiu determinat atât de acuratețea erorii de predicție, cât și de preciziile erorii de predicție. Putem descrie acest lucru într-un mod mai simplu, dacă considerăm că acuratețea erorii de predicție crește odată cu amplitudinea inversă a erorii de predicție în sine, precum și că precizia erorii de predicție crește odată cu amplitudinea inversă a fluctuațiilor aleatorii legate de incertitudinea predicțiilor [...]. Acest lucru justifică o descriere unitară a relației dintre percepția conștientă și atenție. Relația dintre ele este ca între o inferență statistică de gradul întâi și una de gradul doi". Vezi și Hohwy (2012), discutat pe scurt mai sus.

32. Această secțiune se bazează masiv pe Clark (2015), căruia îi sunt recunoscător.

33. Diferența se învârte în jurul faptului că visele aproape că nu conțin semnale de eroare exteroceptive, datorită schimbărilor dramatice la nivelul cântăririi preciziei care apar odată cu instalarea somnului (Honson și Friston, 2012, 2014). Clark descrie acest lucru drept „izolarea față de divertisment", dar vezi nota 35, mai jos.

34. Clark (2015), p. 273.

35. Totuși, observă că creierul în starea de somn nu este „în izolare față de recreere", după cum afirmă Clark (2015), p. 107. Principalul „semnal senzorial de impulsionare" al creierului este întotdeauna endogen. Pur și simplu, acest semnal nu poate să te izoleze de divertisment – adică de nevoile biologice. Dacă noi, specialiștii în neuroștiințe,

continuăm să trecem cu vederea acest fapt fundamental, nu vom înțelege niciodată viața psihică și locul ei în natură.

36. Vezi Domhoff (2017).
37. Clark (2015), p. 274.
38. Vezi Solms (2020a) pentru o dezvoltare a acestei idei.
39. Asta se întâmplă și în „starea de repaus” a modului implicit.
40. E ușor să uităm că hipocampul face parte din sistemul limbic – creierul emoțional.
41. Okuda *et al.* (2003), Szpunar *et al.* (2007), Szpunar (2010), Addis *et al.* (2007).
42. Ingvar (1985).
43. Schacter, Addis și Buckner (2007), p. 660, sublinierea mea.
44. Faptul că un flux *schimbător* (impredictibil) suprimă o imagine *constantă* (mai predictibilă) este, desigur, interesant în sine.
45. Lupyan și Ward (2013), p. 14196, sublinierea mea. În termenii aspectelor tehnice referitoare la percepția conștientă discutate mai sus, acesta este, desigur, un exemplu de atenție endogenă. În acest sens, vezi și definiția tehnică a „predominanței” de la p. 214. Cuvintele stimulează predominanța în mod artificial. Totuși, amorsajul verbal este anulat cu ușurință de o surpriză exogenă puternică (adică precisă). Cu alte cuvinte, prioritizarea de jos în sus a nevoilor la nivelul triunghiului de decizie din mezencefal domină în mod invariabil procesele de sus în jos din prozencefal.
46. Lupyan și Thompson-Schill (2012).
47. Çukur *et al.* (2013). Vezi din nou experimentul despre orbirea perceptivă citat în capitolul 9, nota 8.
48. Clark (2015), p. 286.

49. Roepstorff și Frith (2004).
50. Vezi Zhou *et al.* (2017).
51. Panksepp și Biven (2012), p. 396: „Inițial, *experiențele* audiovizuale conștiente au fost în mare parte afective (Panksepp, 1998). Rapiditatea cu care stimulii vizuali și auditivi subiți pot să ne surprindă și să ne sperie, mai ales când asemenea stimuli își au originea foarte aproape de corpul nostru, sugerează o integrare primară profundă a acestor sisteme senzoriale cu unele dintre mecanismele noastre afective de supraviețuire esențiale. Gândește-te și că suntem înclinați să asociem anumite culori cu sentimente”.
52. Gândește-te și la conceptul tot mai influent al „atingerii afective”.
53. Nu spun că conștiința perceptivă *în totalitatea ei* nu are forță cauzală. Își dobândește forța tocmai deoarece contextualizează afectul.
54. Vezi Hurley, Dennett și Adams (2011), o carte despre umor asupra căreia Dennett mi-a atras atenția recent, deoarece ajunge la concluzii despre arhitectura funcțională a minții remarcabil de similare cu ale mele.
55. Lungimile de undă și intensitățile auditive și vizuale sunt măsurate, comparate și clasificate constant de cortex, atât conștient, cât și inconștient.
56. *Cf.* Clark (2015), p. 207: paradigma energiei libere sugerează „nu că trăim ca atare propriile noastre semnale de eroare în ce privește predicția (sau preciziile lor aferente). În schimb, aceste semnale acționează în noi pentru a coopta fluxurile de predicții adecvate ce dezvăluie o lume de obiecte și cauze distale”.
57. Dacă ești alarmat de faptul că folosesc cuvântul „avatar”,

amintește-ți că tot ce percepi este virtual, inclusiv imaginea propriului corp. Ia în considerare ce am scris despre iluzia „schimbului de corpuri”, în Solms (2013), p. 15: „Subiectul conștiinței se identifică cu corpul său extern (prezentarea obiectului) cam tot așa cum un copil se proiectează pe sine în personajul animat pe care îl controlează într-un joc video. Reprezentarea este învestită rapid cu un simţ al sinelui, deși nu e chiar sinele. Iată un experiment frapant ce ilustrează grăitor relația ilogică care există de fapt între sinele subiectiv și corpul său extern. Petkova și Ehrsson (2008) prezintă o serie de experimente de tip «schimb de corpuri», în cadrul cărora în dreptul ochilor altor oameni sau ai unor manechine s-au amplasat camere video ce au transmis imagini din perspectivele respective către ochelari cu monitorizare video purtați de subiecții experimentului, ceea ce le-a creat rapid subiecților iluzia că corpul celeilalte persoane sau al manechinului e propriul lor corp. Iluzia a fost atât de convingătoare, încât a persistat chiar și atunci când subiecții proiectați au dat mâna cu propriul corp. Existența iluziei a fost demonstrată obiectiv și de faptul că, atunci când celălalt corp (propriu în mod iluzoriu) și propriul corp (real) au fost amenințate cu un cuțit, răspunsul de frică – „reacția instinctivă” a corpului intern (măsurată în funcție de frecvența cardiacă și de conductanța cutanată) – a fost mai puternic în cazul corpului iluzoriu […]. Ne amintim astfel că cortexul nu e decât memorie cu acces aleatoriu”. Pentru această din urmă idee, vezi Ellis și Solms (2018).

Capitolul 11. Problema dificilă

1. Davies (2019), pp. 184, 207.
2. Crick (1994). Citatul complet este dat la p. 249.
3. Chalmers (1995a), p. 201, sublinierea mea.
4. Crick (1994), p. 3, sublinierea mea.
5. Vezi Chalmers (1996), p. 251, sublinierea mea: „Cine ar fi crezut că bucata aceasta de materie cenușie este ceea ce ar putea *să producă* experiențe subiective intense? Și totuși, le produce”.
6. Searle (1997), p. 28. Și Searle se exprimă la fel: „Cum, mai exact, *cauzează* conștiința procesele neurobiologice din creier?” (1993, p. 3, sublinierea mea).
7. Mă refer aici la închiderea cauzală a ceea ce e fizic: dacă conștiința nu este fizică sau dacă proprietățile conștiente nu sunt proprietăți fizice, atunci este greu de înțeles cum pot ele să influențeze matricea cauzală a proceselor cerebrale. Vezi corespondența dintre René Descartes și prințesa Elisabeta a Boemiei (Shapiro, 2007).
8. Levine (1983). Ireductibilitatea experienței fenomenologice la niște procese fizice este numită și „lacuna epistemică”.
9. Jackson (1982). Spun „cam așa” pentru că am modificat puțin „argumentul cunoașterii” adus de Jackson, nu numai ca să-l simplific, ci și pentru că este inutil de crud în forma originală. În lumea reală, o asemenea cruzime ar afecta procesele psihologice descrise de Jackson.
10. Chalmers (2003), p. 104.
11. Vezi *Eseu asupra intelectului omenesc* de Locke (1690): „Este imposibil să concepem că materia, cu sau fără mișcare, a putut de la origine să aibă de la sine simțuri, percepere și

cunoaștere, de unde ar urma deci în chip evident că simțurile, perceperea și cunoașterea sunt însușiri veșnic inseparabile de materie și de fiecare particulă a ei" [John Locke, *Eseu asupra intelectului omenesc*, vol. II, traducere de Teodor Voiculescu, note de Dan Bădărău. București: Editura Științifică, 1961, cartea IV, cap. X, §10, p. 234].

12. Nu vreau să sugerez că Locke a fost un epifenomenalist. Există, desigur, alte poziții dualiste, dar Jackson (1982), care a inventat exercițiul de gândire despre Mary, a adoptat o poziție epifenomenalistă. Mai târziu (1995) s-a răzgândit.

13. Oakley și Halligan (2017). Acești autori atribuie, cel puțin, *o anumită* funcție conștiinței: capacitatea de a *informa despre* stările psihice (care, în sine, sunt inconștiente).

14. Ca să-l parafrazez pe Chalmers, e imaginabil din punct de vedere logic că „zombii filosofici" ar putea să emuleze toate funcțiile mecanice ale creierului fără să aibă vreo experiență conștientă.

15. Libet *et al.* (1983). Totuși, Libet nu crede că conștiința este epifenomenologică. El consideră că, în timpul celor 300 de milisecunde de dinaintea unei acțiuni, conștiința ar putea să aleagă să abandoneze acțiunea respectivă (*free won't*). Acest tip de undă cerebrală se numește „potențial premotor". Studiile ulterioare sugerează că perioada de latență dintre potențialul premotor și decizia conștientă poate fi semnificativ mai mare de 300 de milisecunde.

16. Iar acest lucru poate fi explicat prin legea lui Hebb, care, din același motiv, explică și faptul că experiența ta trăită a amintirilor vechi de zece minute nu este atât de bine consolidată. Trebuie să precizez că toate acestea sunt în concordanță cu ceea ce Chalmers (1995a) numește principiul

coerenței structurale; vezi pp. 262-263.

17. Atkinson și Shiffrin (1971).

18. Urmele mnezice pe termen scurt se degradează rapid din cauza mecanismelor de recaptare a neurotransmițătorilor ce readuc neuronii presinaptici în starea existentă înainte de formarea fiecărei urme, ceea ce le permite imediat să formeze alte urme. Vezi Mongillo, Barak și Tsodyks (2008).

19. Chalmers (1995a), pp. 202-203. Chalmers explică sensul în care folosește termenul „funcție”: „Aici, «funcție» nu este folosit în sensul teleologic limitat, de ceva ce un sistem este conceput să facă, ci în sensul mai larg, de orice rol cauzal în generarea comportamentului pe care l-ar putea manifesta un sistem”.

20. *Ibidem*, pp. 204-205.

21. *Ibidem*, p. 205.

22. *Ibidem*, p. 204.

23. *Cf.* Nagel (1974): „Ca să apărăm fizicalismul, trăsăturile fenomenologice [ale conștiinței] trebuie să aibă o descriere fizică. Dar, atunci când le analizăm caracterul subiectiv, pare că un asemenea rezultat e imposibil. Motivul e că fiecare fenomen subiectiv este conectat, practic, cu un singur punct de vedere și pare inevitabil că o teorie fizică, obiectivă va abandona acest punct de vedere”. Searle (1997), p. 212, spune așa: „Conștiința are o ontologie subiectivă sau la persoana întâi, așa că nu poate fi redusă la ceva ce are ontologie subiectivă sau la persoana a treia. Dacă încerci să o reduci sau să o elimini pe una în favoarea celeilalte, vei lăsa ceva pe dinafară [...]. Nu poți să reduci nici descărcarea neuronilor la sentimente, nici sentimentele la descărcarea neuronilor, deoarece în fiecare caz ar rămâne pe dinafară

obiectivitatea sau subiectivitatea aflată în ecuație”.

24. Chalmers (1995a), p. 203.

25. Havlík, Kozáková și Horáče (2017).

26. Scrisoare din 23 decembrie 2017. Ne-am trimis articolul la *Journal of Consciousness Studies* pentru că acolo a publicat Chalmers (1995a) pentru prima dată problema dificilă.

27. Zahavi (2017) face această afirmație – este o ironie că o face împotriva afirmațiilor lui Friston.

28. Cunosc foarte bine obiecțiile lui Searle în această privință, în sensul că qualia sunt o situație în care o singură realitate nu poate avea mai multe apariții, deoarece ele *sunt* aparițiile (vezi nota 39, mai jos). Răspunsul meu pentru el ar fi: „Da, dar nu uita că apariția *vizualizată* a activității corticale somato-senzoriale și apariția *simțită* a durerii fac parte din aceeași realitate resimțită”.

29. În această analogie, mă refer la două fenomene *exteroceptive*: percepția fulgerului și a tunetului. Când întreb ce le cauzează, aș putea să abordez întrebarea în două feluri. Aș putea fie să descriu evenimentele geofizice (electrice) care le generează, fie mecanismele senzoriale care consemnează (aceste două aspecte diferite legate de) evenimentele fizice. Am ales prima variantă în această analogie, lăsând a doua descriere pentru analiza problemei pe care o tratez în cartea de față, şi anume relația dintre evenimentele obiective și cele subiective. Asta pentru că, după cum am văzut, *conștiința* atașată evenimentelor senzoriale, indiferent dacă sunt activate exteroceptiv sau interoceptiv, este întotdeauna endogenă. Această observație nu poate fi demonstrată în cadrul analogiei. Conștiința în sine, spun eu, nu este un semnal senzorial (exteroceptiv sau

interoceptiv), ci *sentimentul* semnalului.

30. În acest sens, te poți folosi de optogenetică, de exemplu. Utilizând echipamente diferite, poți *să asculți* aşa-numitele *spike trains* ale semnalelor retiniene. După cum spune Wheeler, totul ține doar de „răspunsurile evocate de echipament”.
31. Chalmers (1995a), pentru acest citat și pentru următoarele, sublinierea mea *passim*.
32. La prima vedere, poziția lui Chalmers pare să fie dualism al proprietăților clasic, în care mintea este o proprietate a materiei. Apoi, el concepe atât mintea, cât și materia ca proprietăți a altceva, numit „informație” (vezi mai jos). Asta ar putea să sune a monism al aspectului dual, dar al lui este tipul ciudat de informație pe care l-am discutat în capitolele anterioare: mintea este inerentă informației, și nu receptorului ei. La fel și în cazul aspectului ce ține de materie al informației, după cum consideră Chalmers. Asta înseamnă că cele două aspecte ale informației nu sunt proprietăți în sensul epistemologic (nu sunt *aparițiile* a ceva numit informație), ci în sensul ontologic (sunt aspecte ale informației *înseși*). Totuși, nu asta trebuie să ne preocupe. Ce mă preocupă cel mai mult este afirmația lui Chalmers că *orice* informație are un aspect psihic (ba chiar conștient). Așadar, problema principală nu e dacă el consideră că aspectul psihic al informației este o proprietate sau o substanță. Problema principală devine: este plauzibil să atribuim conștiință oricărei informații?
33. Nu uita că, pentru Chalmers, „caracteristicile fenomenologice ale lumii” exclud „ceea ce este fizic”.
34. Ar trebui să spun: echipamentul *care este* observatorul

participant.

35. Am preluat această analogie din Hurley, Dennett și Adams (2011).

36. Asta are legătură cu ce am spus mai sus (în capitolul 9) despre ipseitatea intențională.

37. Chalmers (1995a), p. 217. Vezi și afirmația lui anterioară: „Asta duce la o ipoteză firească – informația (*sau cel puțin unele informații*) are două aspecte de bază, unul fizic și unul fenomenologic” (sublinierea mea).

38. Vezi capitolul 10, nota 14.

39. *Cf.* Searle (1992), p. 122: „Nu putem [distinge aparența de realitate] în cazul conștiinței, deoarece conștiința constă din aparențe. Când e vorba de aparență, nu putem face distincția între aparență și realitate pentru că aparența este realitatea”.

40. De asemenea, dacă Mary ar fi specialistă în corelativele neurale ale afectului, și nu în cele ale vederii, nu ar fi posibil să nu aibă experiențe afective tot așa cum nu are experiențe vizuale. Asta pentru că, dacă nu ar avea senzații, ar fi în comă (dacă nu moartă).

41. Vezi capitolul 5, nota 4.

42. Ne putem închipui deplasarea cu ajutorul unor roți, dar ceea ce s-a dezvoltat de fapt (în cazul nostru) au fost picioarele. Trebuie să fim atenți să nu ridicăm ștacheta pentru explicațiile referitoare la conștiință mai mult decât cele referitoare la orice altceva din biologie.

43. Chalmers (1995a), pp. 203-204.

Capitolul 12. Cum se face o minte

1. Dar o caracatiță s-ar putea să nu fie de acord. (Sistemul lobar vertical pare să fie cel mai apropiat corespondent la cefalopode al paliumului vertebratelor.)
2. La început, din grup am făcut parte Tristan Hromnik, Jonathan Shock și eu, apoi grupul a inclus şi alți fizicieni, informaticieni și specialiști în bioinginerie medicală – George Ellis, Rowan Hodson, Leen Remmelzwaal, Amit Mishra, Dean Rance, Dawie van den Heever și Julianne Blignaut –, precum și pe neuropsihologii Joshua Martin, Aimee Dollman și Donne van der Westhuizen. Echipa continuă să crească, deși Hromnik nu mai face parte din ea, iar Martin s-a mutat la Berlin.
3. Searle (1980). Damasio (2018) a avut o opinie similară.
4. Mă refer la argumentul „qualia dansante” al lui Chalmers, ce are ca punct de plecare argumentul „spectrelor inversate” al lui Locke. Vezi Chalmers (1995a, 1995b, 2011).
5. Chalmers (1995a), pp. 214-215.
6. *Ibidem*, p. 215, sublinierea mea.
7. *Ibidem*, sublinierea mea.
8. Am avut lungi discuții cu el în această privință și impresia mea e că măcar are o atitudine deschisă.
9. Ethier *et al.* (2012), Hochberg *et al.* (2012), Collinger *et al.* (2013), Bouton *et al.* (2016), Capogrosso *et al.* (2016).
10. Capogrosso *et al.* (2016), p. 284.
11. *Ibidem*.
12. Abu-Hassan *et al.* (2019).
13. Pasley *et al.* (2012).
14. Nishimoto *et al.* (2011).
15. Horikawa *et al.* (2013).
16. Herff *et al.* (2015).

17. Inclusiv cele de sus în jos, de tipul discutat în cadrul temei „gândire” din capitolul 10 (precum imaginarea și visul).
18. *Cf.* Kurzweil (2005).
19. Solms și Turnbull (2002), pp. 70-71, a doua subliniere îmi aparține.
20. Vezi Solms (1996, 1997b).
21. Renumitul fizician Richard Feynman a avut aceeași perspectivă asupra înțelegerii mecaniciste în general. Nu vreau să spun că conștiința obținută prin inginerie inversă rezolvă problema. E posibil să montezi ceva microcip cu microcip fără să te gândești la asta și fără să înțelegi cevaul. Vreau să spun însă că, *dacă* îl înțelegi, atunci ar trebui să-l poți crea prin inginerie inversă.
22. Vezi Reggia (2013), pentru o trecere în revistă a cercetărilor anterioare în acest sens.
23. Întrucât abordarea învățării prin consolidare presupune un criteriu de obiectiv, acesta va fi: *supraviețuirea sistemului în medii neprevăzute*.
24. Ca să fie clar: întruparea poate fi simulată (ca și toți parametrii fizici descriși mai jos). Din punctul de vedere al sistemului, nu contează ce se întâmplă de fapt „în exterior”, ci numai ce se întâmplă în model în legătură cu *informațiile pe care le primește* din exterior. Prin urmare, echipa de cercetare poate să simuleze un mediu din care sistemul să-și facă un model și vom proceda astfel pentru primele generații ale sistemului pe care îl propunem. Dacă am proceda altfel, ar fi nevoie de extrem de mult timp și, sincer, ar fi periculos (de exemplu, gândește-te la parametrul de supraîncălzire, descris mai jos). Când spun că este mai „realist” să întrupăm sistemul din punct de vedere fizic, mă

refer la faptul că, odată cu mișcarea fizică (de exemplu), apar probleme de modelare ce nu apar în cazul mișcării simulate, iar acest lucru s-ar putea dovedi a fi important pentru un sistem cu adevărat similar unuia însuflețit. Din acest motiv și din altele, generațiile ulterioare ale sistemului pe care îl propunem vor fi întrupate în roboți.

25. Adică vom codifica reflexe și instincte artificiale. Vom folosi și „algoritmi genetici”.

26. De pe Wikipedia (https://en.wikipedia.org/wiki/Artificial_consciousness), 21 martie 2020. Pentru o opinie alternativă, vezi Reggia (2013): „Autorul acestei recenzii crede că niciunul dintre studiile anterioare examinate, chiar dacă se afirmă altceva, nu oferă încă vreun argument convingător cu privire la modul cum va duce abordarea studiată în cele din urmă la o conștiință artificială concretizată. Totodată și pe un ton mai marcat pozitiv, încă nu s-au adus dovezi (inclusiv în studiile analizate în această recenzie) că, în cele din urmă, conștiința automată concretizată nu ar fi posibilă, opinie exprimată și de alții”.

27. Solms și Turnbull (2002), pp. 68-69.

28. Vezi Colby, Watt și Gilbert (1966), Weizenbaum (1976). Un program de calculator numit „Eugene Goostman”, care simulează un băiat ucrainean de 13 ani, a trecut testul Turing în cadrul unui eveniment organizat în 2014 la Royal Society of London.

29. De exemplu, după cum se știe, calculatoarele îi pot bate și pe cei mai buni jucători umani și de șah, și de go, care este mai dificil decât șahul.

30. Vezi Haikonen (2012).

31. Este emoționant, având în vedere cât a suferit chiar Alan

Turing (creatorul testului) din cauza prejudecăților.

32. Acum mai mulți ani, am spus în cadrul unei reuniuni desfășurate la Viena între psihanaliști și ingineri specialiști în inteligență artificială că existența conștiinței artificiale ar putea fi demonstrată, de exemplu, dacă s-ar găsi dovezi de psihopatologie artificială: „În măsura în care inginerii reușesc să imite cu acuratețe mintea umană, vor descoperi că modelul lor este predispus la anumite tipuri de disfuncționalitate. Aproape că suntem tentați să folosim acest lucru drept criteriu al succesului lor” (Solms, 2008).
33. Mathur, Lau și Guo (2011).
34. Gândește-te la cazul extrem al „corporațiilor autonome descentralizate”.
35. Vezi Lin, Abney și Bekey (2011).
36. Bill Gates, Stephen Hawking și Elon Musk, de exemplu, și-au exprimat rezerve serioase pe această temă.
37. Nu pretind că această convingere, singură, constituie o justificare etică pentru un asemenea demers. Faptul că cineva va comite o crimă undeva, la un moment dat, nu justifică o crimă a mea aici și acum. Te rog să citești mai departe…
38. Solms și Friston (2018), Solms (2019a), Solms (2020b).
39. Iată o listă parțială: „Where does consciousness fit in the Bayesian brain?”, al 18-lea Congres Internațional de Neuropsihanaliză, University College London, 2017; „How and why consciousness arises”, departamentul de fizică, University of Cape Town, 2017; „How and why consciousness arises”, Centrul de studii asupra subiectivității, Københavns Universitet, Copenhaga, 2017; „The self as feeling and memory”, Ruhr Universität, Bochum,

2018; „The conscious id, the psychoanalytic process and the hard problem of consciousness”, departamentul de filosofie, New York University, 2019; „Why and how consciousness arises”, departamentul de psihiatrie, spitalul Mount Sinai, New York, 2019; „Why are we conscious? Lessons from neuroscience”, University of Vermont College of Medicine, Burlington, 2019; „What is consciousness?”, The Melbourne Brain Centre, Australia, 2019; „Consciousness itself”, The Science of Consciousness, Interlaken, Elveția, 2019; „Consciousness itself is affect”, Hochschule für Philosophie, München, Burkardus Haus, Würzburg, Germania, 2019; „The hard problem of consciousness”, departamentul de filosofie, University of Cape Town, 2019; „Consciousness is predictive work in progress”, spitalul Ichilov, Tel Aviv, 2019; „Why and how consciousness arises”, Italian Psychoanalytic Dialogues, Roma, 2020.

40. Havlík, Kozáková și Horáče (2017).

41. Carhart-Harris și Friston (2010), Carhart-Harris *et al.* (2014), Carhart-Harris (2018).

42. Vezi capitolul 9, nota 16.

43. www.linking-ai-principles.org/term/656. Vezi The Public Voice Coalition (2018).

Post-scriptum

1. Legea afectului: „Dacă un comportament este însoțit constant de plăcere, se va intensifica, iar dacă este însoțit constat de neplăcere, va scădea”.

2. Legea lui Friston: „Toate cantitățile care se pot schimba, cu

alte cuvinte, care fac parte din sistem, se vor schimba cu siguranţă pentru a minimiza energia liberă”.

Bibliografie

Abbott, A. (2020), What animals really think. *Nature*, 584: 182-185.

Absher, J. și Benson, D. (1993), Disconnection syndromes: an overview of Geschwind's contributions. *Neurology*, 43: 862-867.

Abu-Hassan, K., Taylor, J., Morris, P. *et al.* (2019), Optimal solid state neurons. *Nature Communications*, 10: 5309.

Adams, R., Shipp, S. și Friston, K. (2013), Predictions not commands: active inference in the motor system. *Brain Structure and Function*, 218: 611-643.

Addis, D., Wong, A. și Schacter, D. (2007), Remembering the past and imagining the future: common and distinct neural substrates during event construction and elaboration. *Neuropsychologia*, 45: 1363-1377.

Ainley, V., Apps, M.A.J., Fotopoulou, A. și Tsakiris, M. (2016), „Bodily precision": a predictive coding account of individual differences in interoceptive accuracy. *Philosophical Transactions of the Royal Society of London, B*, 371: (1708), doi.org/10.1098/rstb.2016.0003

Alboni, P. (2014), Vasovagal syncope as a manifestation of an evolutionary selected trait. *Journal of Atrial Fibrillation*, 7: 1035.

Aserinsky, E. și Kleitman, N. (1953), Regularly occurring periods of eye motility, and concomitant phenomena, during sleep. *Science*, 118: 273-274.

Ashby, W. (1947), Principles of the self-organizing dynamic system. *Journal of General Psychology*, 37: 125-128.

Atkinson, R. și Shiffrin, R. (1971), The control of short-term memory. *Scientific American*, 225: 82-90.

Baars, B. (1989), *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press.

Baars, B. (1997), *In the Theatre of Consciousness*. Oxford: Oxford University Press.

Bailey, P. și Davis, E. (1942), The syndrome of obstinate progression in the cat. *Experimental Biology and Medicine*, 52: 307.

Bargh, J. și Chartrand, T. (1999), The unbearable automaticity of being. *American Psychologist*, 54: 462-479.

Barrett, L.F. (2017), *How Emotions are Made: The Secret Life of the Brain*. New York: Houghton Mifflin Harcourt.

Bastos, A., Usrey, W., Adams, R. *et al.* (2012), Canonical microcircuits for predictive coding. *Neuron*, 76: 695-711.

Bastos, A., Vezoli, J., Bosman, C. *et al.* (2015), Visual areas exert feedforward and feedback influences through distinct frequency channels. *Neuron*, 85: 390-401.

Bayes, T. (1763), An essay towards solving a problem in the doctrine of chances. [Communicated by Mr. Price, in a letter to John Canton.] *Philosophical Transactions of the Royal Society of London*, 53: 370-418.

Bechtel, W. și Richardson, R. (1998), Vitalism. În E. Craig (ed.), *Routledge Encyclopedia of Philosophy*, 9. Londra: Routledge, pp. 639-643.

Bentley, B., Branicky, R., Barnes, C. *et al.* (2016), The multilayer connectome of *Caenorhabditis elegans*. *PLoS Computational Biology*, 12: e1005283, doi.org/10.1371/journal.pcbi.1005283

Berlin, H. (2011), The neural basis of the dynamic unconscious. *Neuropsychoanalysis*, 13: 5-31.

Berlin, H. (2013), The brainstem begs the question: „petitio principii". *Neuropsychoanalysis*, 15: 25-29.

Berridge, K. (2003), Pleasures of the brain. *Brain and Cognition*, 52: 106-128.

Besharati, S., Forkel, S.J., Kopelman, M., Solms, M., Jenkinson, P.M. și Fotopoulou, A. (2014), The affective modulation of motor awareness in anosognosia for hemiplegia: behavioural and lesion evidence. *Cortex*, 61: 127-140.

Besharati, S., Forkel, S., Kopelman, M., Solms, M., Jenkinson, P. și Fotopoulou, A. (2016), Mentalizing the body: spatial and social cognition in anosognosia for hemiplegia. *Brain*, 139: 971-985.

Besharati, S., Fotopoulou, A. și Kopelman, M. (2014), What is it like to be confabulating? În A.L. Mishara, A. Kranjec, P. Corlett, P. Fletcher și M.A. Schwartz (eds.), *Phenomenological Neuropsychiatry: How Patient Experience Bridges Clinic with Clinical Neuroscience*. New York: Springer.

Bienenstock, E., Cooper L. și Munro P. (1982), Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *Journal of Neuroscience*, 2: 32-48.

Blake, Y., Terburg, D., Balchin, R., van Honk, J. și Solms, M. (2019), The role of the basolateral amygdala in dreaming, *Cortex*, 113: 169-183, doi.org/10.1016/j.cortex.2018.12.016

Block, N. (1995), On a confusion about a function of

consciousness. *Behavioral and Brain Sciences*, 18: 227-287.
Blomstedt, P., Hariz, M., Lees, A. *et al.* (2008), Acute severe depression induced by intraoperative stimulation of the substantia nigra: a case report. *Parkinsonism and Related Disorders*, 14: 253-156.
Bogen, J. (1995), On the neurophysiology of consciousness: 1. An overview. *Consciousness and Cognition*, 4: 52-62.
Bouton, C., Shaikhouni, A., Annetta, N. *et al.* (2016), Restoring cortical control of functional movement in a human with quadriplegia. *Nature*, 533: 247-250.
Bowlby, J. (1969), *Attachment*. Londra: Hogarth Press.
Braun, A. (1999), The new neuropsychology of sleep. *Neuropsychoanalysis*, 2: 196-201.
Braun, A., Balkin, T., Wesenten, N. *et al.* (1997), Regional cerebral blood flow throughout the sleep-wake cycle. An H2(15)O PET study. *Brain*, 120: 1173-1197.
Brentano, F. (1874), *Psychologie vom empirischen Standpunkte*. Leipzig: Duncker and Humbolt.
Broca, P. (1861), Sur le principe des localisations cerebrales. *Bulletin de la Societe d'Anthropologie*, 2: 190-204.
Broca, P. (1865), Sur le siege de la faculte du langage articule. *Bulletin de la Societe d'Anthropologie*, 6: 377-373.
Brown, H., Adams, R., Parees, I., Edwards, M. și Friston, K. (2013), Active inference, sensory attenuation and illusions. *Cognitive Processing*, 14: 411-427.
Cameron-Dow, C. (2012), Do dreams protect sleep? Testing the Freudian hypothesis of the function of dreams. MA dissertation, University of Cape Town.
Campbell, A. (1904), Histological studies on the localisation of cerebral function. *Journal of Mental Science*, 50: 651-659.

Capogrosso, M., Milekovic, T., Borton, D. *et al.* (2016), A brain-spine interface alleviating gait deficits after spinal cord injury in primates. *Nature*, 539: 284-288.

Carhart-Harris, R. (2018), The entropic brain – revisited. *Neuropharmacology*, 142: 167-178.

Carhart-Harris, R. și Friston, K. (2010), The default-mode, ego-functions and free-energy: a neurobiological account of Freudian ideas. *Brain*, 133: 1265-1283.

Carhart-Harris, R., Leech, R., Hellyer, P. *et al.* (2014), The entropic brain: a theory of conscious states informed by neuroimaging research with psychedelic drugs. *Frontiers in Human Neuroscience*, 8: Article 20.

Chabris, C. și Simons, D. (2010), *The Invisible Gorilla: and Other Ways Our Intuitions Deceive Us.* Londra: Crown Publishers/Random House.

Chalmers, D. (1995a), Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 2: 200-219.

Chalmers, D. (1995b), Absent qualia, fading qualia, dancing qualia. În T. Metzinger (ed.), *Conscious Experience*. Paderborn: Ferdinand Schoningh, pp. 309-328.

Chalmers, D. (1996), *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.

Chalmers, D. (2003), Consciousness and its place in nature. În S. Stich și T. Warfield (eds.), *Blackwell Guide to the Philosophy of Mind*. Londra: Blackwell, pp. 102-142.

Chalmers, D. (2011), A computational foundation for the study of cognition. *Journal of Cognitive Science*, 12: 325-359.

Charcot J.-M. (1883), Un cas de suppression brusque et isolee de la vision mentale des signes et des objets (formes et couleurs). *Progres Medical*, 11: 568.

Chew, Y., Tanizawa, Y., Cho, Y. *et al.* (2018), An afferent neuropeptide system transmits mechanosensory signals triggering sensitization and arousal in *C. elegans*. *Neuron,* 99: 1233-1246.

Cisek, P. și Kalaska, J. (2010), Neural mechanisms for interacting with a world full of action choices. *Annual Review of Neuroscience*, 33: 269-298.

Claparede, E. (1911), Recognition et moitié. *Archives of Psychology Geneve*, 11: 79-90.

Clark, A. (2015), *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. New York: Oxford University Press.

Clark, A. (2017), Busting out: predictive brains, embodied minds, and the puzzle of the evidentiary veil. *Nous*, 51: 727-753.

Coenen, A. (2007), Consciousness without a cortex, but what kind of consciousness is this? *Behavioral and Brain Sciences*, 30: 87-88.

Coenen, V., Bewernick, B., Kayser, S. *et al.* (2019), Superolateral medial forebrain bundle deep brain stimulation in major depression: a gateway trial. *Neuropsychopharmacology*, 44: 1224-1232, doi.org/ 10.1038/s41386–019–0369–9

Colby, K., Watt, J. și Gilbert, J. (1966), A computer method of psychotherapy. *Journal of Nervous and Mental Disease*, 142: 148-152.

Cole, S., Fotopoulou, A., Oddy, M. și Moulin, C. (2014), Implausible future events in a confabulating patient with an anterior communicating artery aneurysm. *Neurocase*, 20: 208-224.

Collinger J., Wodlinger, B., Downey, J. *et al.* (2013), High-performance neuroprosthetic control by an individual with

tetraplegia. *The Lancet*, 381: 557-564.
Coltheart, M. și Turner, M. (2009), Confabulation and delusion. În W. Hirstein (ed.), *Confabulation: Views from Neuroscience, Psychiatry, Psychology and Philosophy*. New York: Oxford University Press, p. 173.
Conant, R. și Ashby, W. (1970), Every good regulator of a system must be a model of that system. *International Journal of Systems Science*, 1: 89-97.
Coren, S. și Porac, C. (1974), The fading of stabilized images: eye movements and information processing. *Perception & Psychophysics*, 16: 529-534.
Corlett, P. și Fletcher, P. (2014), Computational psychiatry: a Rosetta Stone linking the brain to mental illness. *Lancet Psychiatry*, 1: 399-402.
Craig, A.D. (2009), How do you feel – now? The anterior insula and human awareness. *Nature Reviews Neuroscience*, 10: 59-70.
Craig, A.D. (2011), Significance of the insula for the evolution of human awareness of feelings from the body. *Annals of the New York Academy of Sciences*, 1225: 72-82.
Crick, F. (1994), *The Astonishing Hypothesis: The Scientific Search for the Soul*. New York: Charles Scribner's Sons.
Crick, F. (2004), Foreword. În C. Koch, *The Quest for Consciousness: A Neurobiological Approach*. Englewood, CO: Roberts and Company.
Crick, F. și Koch, C. (1990), Towards a neurobiological theory of consciousness. *Seminars in Neuroscience*, 2: 263-275.
Crucianelli, L., Krahe, C., Jenkinson, P. și Fotopoulou, A. (2017), Interoceptive ingredients of body ownership: affective touch and cardiac awareness in the rubber hand illusion. *Cortex*,

104: 180-192, doi.org/10.1016/j.cortex.2017.04.018
Cukur, T., Nishimoto, S., Huth, A. și Gallant J. (2013), Attention during natural vision warps semantic representation across the human brain. *Nature Neuroscience*, 16: 763-770.
Dahan, L., Astier, B., Vautrelle, N. *et al.* (2007), Prominent burst firing of dopaminergic neurons in the ventral tegmental area during paradoxical sleep. *Neuropsychopharmacology*, 32: 1232-1241.
Damasio, A. (1994), *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Putnam.
Damasio, A. (2018), *The Strange Order of Things: Life, Feeling, and the Making of Cultures.* Londra: Penguin Random House.
Damasio, A. și Carvalho, G. (2013), The nature of feelings: evolutionary and neurobiological origins. *Nature Reviews Neuroscience*, 14: 143-152.
Damasio, A. și Damasio, H. (1989), *Lesion Analysis in Neuropsychology*. New York: Oxford University Press.
Damasio, A., Damasio, H. și Tranel, D. (2013), Persistence of feelings and sentience after bilateral damage of the insula. *Cerebral Cortex*, 23: 833-846.
Damasio, A., Grabowski, T., Bechara, A. *et al.* (2000), Subcortical and cortical brain activity during the feeling of self-generated emotions. *Nature Neuroscience*, 3: 1049-1056.
Darwin, C. (1859), *On the Origin of Species*. Londra: John Murray.
Darwin, C. (1872), *The Expression of Emotions in Man and Animals*. Londra: John Murray.
Davies, P. (2019), *The Demon in the Machine: How Hidden Webs of Information are Solving the Mystery of Life.* Londra: Allen Lane.

Debiec, J., Doyere, V., Nader, K. și LeDoux, J. (2006), Directly reactivated, but not indirectly reactivated, memories undergo reconsolidation in the amygdala. *Proceedings of the National Academy of Sciences*, 103: 3428-3433.

Decety, J. și Fotopoulou, A. (2015), Why empathy has a beneficial impact on others in medicine: unifying theories. *Frontiers in Behavioral Neuroscience*, 8: 457.

Dehaene, S. și Changeux, J.-P. (2005), Ongoing spontaneous activity controls access to consciousness: a neuronal model for inattentional blindness. *PLoS Biology*, 3: e141.

Dehaene, S. și Changeux, J.-P. (2011), Experimental and theoretical approaches to conscious processing. *Neuron*, 70: 200-227.

Dehaene, S. și Naccache, L. (2001), Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition*, 79: 1-37.

Dement, W. și Kleitman, N. (1957), The relation of eye movements during sleep to dream activity: an objective method for the study of dreaming. *Journal of Experimental Psychology*, 53: 339-346.

Depaulis, A. și Bandler, R. (1991), *The Midbrain Periaqueductal Gray Matter: Functional, Anatomical, and Neurochemical Organization*. New York: Plenum Press.

Ditchburn, R. și Ginsborg, B. (1952), Vision with a stabilized retinal image, *Nature*, 170: 36-37.

Domhoff, W. (2017), *The Emergence of Dreaming: Mind-Wandering, Embodied Simulation, and the Default Network*. New York: Oxford University Press.

Du Bois-Reymond, E. (1848-1884), *Untersuchungen uber thierische Electricität*, 2. Berlin: Reimer.

Du Bois-Reymond, E., ed. (1918), *Jugendbriefe von Emil Du Bois-Reymond an Eduard Hallmann, zu seinem hundertsten Geburtstag, dem 7. November 1918*. Berlin: Reimer.

Dudai, Y. (2000), The shaky trace. *Nature*, 406: 686-687.

Edelman, G. (1990), *The Remembered Present: A Biological Theory of Consciousness*. New York: Basic Books.

Edlow, B., Takahashi, E., Wu, O. *et al.* (2012), Neuroanatomic connectivity of the human ascending arousal system critical to consciousness and its disorders. *Journal of Neuropathology and Experimental Neurology*, 71: 531-546.

Einstein, A. (1905), Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt. *Annalen der Physik*, 17: 132-148.

Eisenberger, N. (2012), The neural bases of social pain: evidence for shared representations with physical pain. *Psychosomatic Medicine*, 74: 126-135.

Ekman, P., Friesen, W., O'Sullivan, M. *et al.* (1987), Universals and cultural differences in the judgements of facial expressions of emotion. *Journal of Personality and Social Psychology*, 53: 712-717.

Ellis, G. și Solms, M. (2018), *Beyond Evolutionary Psychology: How and Why Neuropsychological Modules Arise*. Cambridge: Cambridge University Press.

England, J. (2013), Statistical physics of self-replication. *Journal of Chemical Physics*, 139: 121923, doi.org/10.1063/1.4818538

Ethier, C., Oby, E., Bauman, M. și Miller, L. (2012), Restoration of grasp following paralysis through brain-controlled stimulation of muscles. *Nature*, 485: 368-371.

Ezra, M., Faull, O., Jbabdi, S. și Pattinson, K. (2015), Connectivity-based segmentation of the periaqueductal gray matter in

human with brainstem optimized diffusion MRI. *Human Brain Mapping*, 36: 3459-3471.
Feldman, H. și Friston, K.J. (2010), Attention, uncertainty, and freeenergy. *Frontiers in Human Neuroscience*, 4: 215, doi.org/10.3389/fnhum.2010.00215
Ferrarelli, F. și Tononi, G. (2011), The thalamic reticular nucleus and schizophrenia. *Schizophrenia Bulletin*, 37: 306-315.
Fischer, D., Boes, A., Demertzi, A. *et al.* (2016), A human brain network derived from coma-causing brainstem lesions. *Neurology*, 87: 2427-2434.
Flechsig, P. (1901), Developmental (mylogenetic) localisation of the cerebral cortex in the human subject. *The Lancet*, 2: 1027-1029.
Flechsig, P. (1905), Gehirnphsyiologie und Willenstheorien. *Fifth International Psychology Congress*, Rome, pp. 73-89. În G. von Bonin (ed.), *Some Papers on the Cerebral Cortex*. Springfield, IL: Charles C. Thomas, pp. 181-200.
Forrester, G., Davis, R., Mareschal, D. *et al.* (2018), The left cradling bias: an evolutionary facilitator of social cognition? *Cortex*, 118: 116-131, doi.org/10.1016/j.cortex.2018.05.011
Fotopoulou, A. (2008), False-selves in neuropsychological rehabilitation: the challenge of confabulation. *Neuropsychological Rehabilitation*, 18: 541-565.
Fotopoulou, A. (2009), Disentangling the motivational theories of confabulation. În W. Histein (ed.), *Confabulation: Views from Neurology, Psychiatry, and Philosophy*. New York: Oxford University Press.
Fotopoulou, A. (2010a), The affective neuropsychology of confabulation and delusion. *Cognitive Neuropsychiatry*, 15: 38-63.

Fotopoulou, A. (2010b), The affective neuropsychology of confabulation and delusion. În R. Langdon și M. Turner (eds.), *Confabulation and Delusion*, New York: Psychology Press, pp. 38-63.

Fotopoulou, A. (2013), Beyond the reward principle: consciousness as precision seeking. *Neuropsychoanalysis*, 15: 33-38.

Fotopoulou, A. și Conway, M. (2004), Confabulation pleasant and unpleasant. *Neuropsychoanalysis*, 6: 26-33.

Fotopoulou, A., Conway, M., Birchall, D., Griffiths, P. și Tyrer, S. (2007), Confabulation: revising the motivational hypothesis. *Neurocase*, 13: 6-15.

Fotopoulou, A., Conway, M. și Solms, M. (2007), Confabulation: motivated reality monitoring. *Neuropsychologia*, 45: 2180-2190.

Fotopoulou, A., Conway, M., Solms, M., Kopelman, M. și Tyrer, S. (2008a), Self-serving confabulation in prose recall. *Neuropsychologia*, 46: 1429-1441.

Fotopoulou, A., Conway, M., Tyrer, S., Birchall, D., Griffiths, P. și Solms, M. (2008b), Is the content of confabulation positive? An experimental study. *Cortex*, 44: 764-772.

Fotopoulou, A., Solms, M. și Turnbull, O. (2004), Wishful reality distortions in confabulation: a case report. *Neuropsychologia*, 42: 727-744.

Fotopoulou, A. și Tsakiris, M. (2017), Mentalizing homeostasis: the social origins of interoceptive inference. *Neuropsychoanalysis*, 19: 3-76.

Frank, J. (1946), Clinical survey and results of 200 cases of prefrontal leucotomy. *Journal of Mental Sciences*, 92: 497-508.

Frank, J. (1950), Some aspects of lobotomy (prefrontal leucotomy) under psychoanalytic scrutiny. *Psychiatry*, 13: 45-52.

Frank, M. (2005), Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. *Journal of Cognitive Neuroscience*, 1: 51-72.

Freud, S. (1883), Einleitung in der Nervenpathologie. Unpublished book manuscript. Washington, DC: Library of Congress.

Freud, S. (1886), Report on my studies in Paris and Berlin. *Standard Edition of the Complete Psychological Works of Sigmund Freud*, 1. Londra: Hogarth, pp. 1-15.

Freud, S. (1888), Gehirn. I. Anatomie des Gehirns. În A. Villaret (ed.), *Handwörterbuch der gesamten Medizin*, 1. Stuttgart: Ferdinand Enke, pp. 684-691.

Freud, S. (1891), *On Aphasia*. New York: International Universities Press.

Freud, S. (1893a), Charcot. *Standard Edition of the Complete Psychological Works of Sigmund Freud*, 3. Londra: Hogarth, pp. 11-23.

Freud, S. (1893b), Some points for a comparative study of organic and hysterical motor paralyses. *Standard Edition of the Complete Psychological Works of Sigmund Freud*, 1. Londra: Hogarth, pp. 155-172.

Freud, S. (1894), The neuro-psychoses of defence. *Standard Edition of the Complete Psychological Works of Sigmund Freud*, 3. Londra: Hogarth, pp. 45-61.

Freud, S. (1895), Studies on hysteria. *Standard Edition of the Complete Psychological Works of Sigmund Freud*, 2. Londra:

Hogarth [ed.rom.: *Studii despre isterie*, traducere din limba germană de Maria şi Ion Nastasia, cuvânt înainte de Vasile Dem. Zamfirescu. Bucureşti: Trei, 2014].

Freud, S. (1900), The interpretation of dreams. *Standard Edition of the Complete Psychological Works of Sigmund Freud*, 4 and 5. Londra: Hogarth [ed.rom.: *Opere esenţiale*, vol. 2: *Interpretarea viselor*, traducere din limba germană de Roxana Melnicu. Bucureşti: Trei, 2010].

Freud, S. (1901), The psychopathology of everyday life. *Standard Edition of the Complete Psychological Works of Sigmund Freud*, 6. Londra: Hogarth [ed.rom.: *Psihopatologia vieţii cotidiene. Despre uitare, greşeli de vorbire, superstiţie şi eroare*, traducere de Herta Spuhn şi Daniela Ştefănescu, ediţia a II-a. Bucureşti: Trei, 2010, e-book].

Freud, S. (1912), A note on the unconscious in psycho-analysis. *Standard Edition of the Complete Psychological Works of Sigmund Freud*, 12. Londra: Hogarth, pp. 255-266.

Freud, S. (1914), On narcissism: an introduction. *Standard Edition of the Complete Psychological Works of Sigmund Freud*, 14. Londra: Hogarth, pp. 67-102 [ed. rom.: „Pentru a introduce narcisismul", traducere de Gilbert Lepădatu, în *Opere esenţiale*, vol. 3: *Psihologia inconştientului*, notă asupra ediţiei de Raluca Hurduc, note introductive de Roxana Melnicu. Bucureşti: Trei, 2017, pp. 35-68].

Freud, S. (1915a), Instincts and their vicissitudes. *Standard Edition of the Complete Psychological Works of Sigmund Freud*, 14. Londra: Hogarth, pp. 117-140 [ed.rom.: „Pulsiuni şi destine ale pulsiunilor", traducere de Gilbert Lepădatu, în *Opere esenţiale*, vol. 3: *Psihologia inconştientului*, notă asupra ediţiei de Raluca Hurduc, note introductive de

Roxana Melnicu. Bucureşti: Trei, 2017, pp. 71-96].
Freud, S. (1915b), The unconscious. *Standard Edition of the Complete Psychological Works of Sigmund Freud*, 14. Londra: Hogarth, pp. 166-204.
Freud, S. (1920), Beyond the pleasure principle. *Standard Edition of the Complete Psychological Works of Sigmund Freud*, 18. Londra: Hogarth, pp. 7-64 [ed.rom.; „Dincolo de principiul plăcerii", în *Opere esenţiale*, vol. 3: *Psihologia inconştientului*, traducere de G. Purdea şi Vasile Dem. Zamfirescu. Bucureşti: Trei, 2017, pp. 213-286].
Freud, S. (1923), The ego and the id. *Standard Edition of the Complete Psychological Works of Sigmund Freud*, 19. Londra: Hogarth, pp. 12-59.
Freud, S. (1925), A note upon „the mystic writing-pad". *Standard Edition of the Complete Psychological Works of Sigmund Freud*, 16. Londra: Hogarth, pp. 227-232.
Freud, S. (1940 [1939]), An outline of psycho-analysis. *Standard Edition of the Complete Psychological Works of Sigmund Freud*, 23. Londra: Hogarth, pp. 144-207 [ed.rom.: „Compendiu de psihanaliză", în *Despre psihanaliză*, traducere de Daniela Ştefănescu, prefaţă de Vasile Dem. Zamfirescu. Bucureşti: Trei, 2014, pp. 169-253].
Freud, S. (1950a [1895]), Extracts from the Fliess papers. *Standard Edition of the Complete Psychological Works of Sigmund Freud*, 1. Londra: Hogarth, pp. 177-280.
Freud, S. (1950b [1895]), Project for a scientific psychology. *Standard Edition of the Complete Psychological Works of Sigmund Freud*, 1. Londra: Hogarth, pp. 283-397 [ed.rom.: *Opere*, vol. XVII: *Proiect de psihologie*, traducere din limba germană de Daniela Ştefănescu, cuvânt înainte de Vasile

Dem. Zamfirescu. Bucureşti: Trei, 2007].
Freud, S. (1994 [1929]), Letter to Einstein, 1929. În I. Grubrich-Simitis (1995), „No greater, richer, more mysterious subject … than the life of the mind”. *International Journal of Psychoanalysis*, 76: 115-122.
Friston, K. (2005), A theory of cortical responses. *Philosophical Transactions of the Royal Society of London, B*, 360: 815-836.
Friston, K. (2009), The free-energy principle: a rough guide to the brain? *Trends in Cognitive Sciences*, 13: 293-301.
Friston, K. (2013), Life as we know it. *Journal of the Royal Society Interface*, 10: 20130475, doi.org/10.1098/rsif.2013.0475
Friston, K., Breakspear, M. și Deco, G. (2012), Perception and selforganized instability. *Frontiers in Computational Neuroscience*, 6: 44.
Friston, K., Rigoli, F., Ognibene, D. *et al.* (2015), Active inference and epistemic value. *Cognitive Neuroscience*, 6: 187-214.
Friston, K., Schwartenbeck, P., FitzGerald, T., Moutoussis, M., Behrens, T. și Dolan, R. (2014), The anatomy of choice: dopamine and decision-making. *Philosophical Transactions of the Royal Society of London, B*, 369: doi.org/10.1098/rstb.2013.0481
Friston, K., Shiner, T., Fitzgerald, T., Galea, J., Adams, R., Brown, H., Dolan, R., Moran, R., Stephan, K. și Bestmann, S. (2012), Dopamine, affordance and active inference. *PLoS Computational Biology*, 8: e1002327
Friston, K. și Stephan, K. (2007), Free-energy and the brain. *Synthese*, 159: 417-458.
Friston, K., Stephan, K., Montague, R. și Dolan, R. (2014), Computational psychiatry: the brain as a phantastic organ. *Lancet Psychiatry*, 1: 148-158.

Frith, C., Blakemore, S. și Wolpert, D. (2000), Abnormalities in the awareness and control of action. *Philosophical Transactions of the Royal Society of London, B*, 355: 1771-1788.

Galin, D. (1974), Implications for psychiatry of left and right cerebral specialization: a neurophysiological context for unconscious processes. *Archives of General Psychiatry*, 31: 572-583.

Garcia-Rill, E. (2017), Bottom-up gamma and stages of waking. *Medical Hypotheses*, 104: 58-62.

Gloor, P. (1992), Role of the amygdala in temporal lobe epilepsy. În J. Aggleton (ed.), *The Amygdala: Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction*. New York: Wiley-Liss, pp. 505-538.

Golaszewski, S. (2016), Coma-causing brainstem lesions. *Neurology*, 87: 10.

Goodglass, H. (1986), Norman Geschwind (1926-1984). *Cortex*, 22: 7-10.

Gosseries, O., Schnakers, C., Ledoux, D. *et al.* (2011), Automated EEG entropy measurements in coma, vegetative state/unresponsive wakefulness syndrome and minimally conscious state. *Functional Neurology*, 26: 25-30.

Gregory, R. (1980), Perceptions as hypotheses. *Philosophical Transactions of the Royal Society of London, B*, 290: 181-197.

Haikonen, P. (2012), *Consciousness and Robot Sentience*. New Jersey: World Scientific.

Harding, D. (1961), *On Having No Head*. Londra: Sholland Trust.

Harlow, J. (1868), Passage of an iron rod through the head. *Boston Medical and Surgical Journal*, 39: 389-393.

Hartmann. E., Russ. D., Oldfield. M., Falke. R. și Skoff. B. (1980),

Dream content: effects of 1-DOPA. *Sleep Research*, 9: 153.
Hassin, R., Bargh, J., Engell, A. și McCulloch, K. (2009), Implicit working memory. *Consciousness and Cognition*, 18: 665-678.
Havlik, M., Kozakova, E. și Horače, J. (2017), Why and how: the future of the central questions of consciousness. *Frontiers in Psychology*, 8: 1797, doi.org/10.3389/fpsyg.2017.01797
Hebb, D. (1949), *The Organization of Behavior: A Neuropsychological Theory*. New York: Wiley.
Helmholtz, H. von (1867), *Handbuch der physiologischen Optik*, 3. Leipzig: Voss.
Helmholtz, H. von (1892), Goethes Vorahnungen kommender naturwissenschaftlicher Ideen. În *Vorträge und Reden*, 2. Braunschweig: Friedrich Vieweg und Sohn, pp. 335-361.
Herff, C., Heger, D., De Pesters, A. *et al.* (2015), Brain-to-text: decoding spoken phrases from phone representations in the brain. *Frontiers in Neuroscience*, 9: 217, doi.org/10.3389/fnins.2015.00217
Hering, E. (1879), Der Raumsinn und die Bewegungen des Auges. În L. Hermann (ed.), *Handbuch der Physiologie*, 3. Part 1: *Physiologie des Gesichtssinnes*. Leipzig: Vogel, pp. 343-601.
Hesselmann, G., Sadaghiani, S., Friston, K. și Kleinschmidt, A. (2010), Predictive coding or evidence accumulation? False inference and neuronal fluctuations. *PLoS One*, 5(3): e9926, doi.org/10.1371/journal.pone.0009926
Hobson, J.A. (2009), REM sleep and dreaming: towards a theory of protoconsciousness. *Nature Reviews Neuroscience*, 10: 803-813.
Hobson, J.A. și Friston (2012), Waking and dreaming consciousness: neurobiological and functional considerations. *Progress in Neurobiology*, 98: 82-98.

Hobson, J.A. și Friston, K. (2014), Consciousness, dreams, and inference: the Cartesian theatre revisited. *Journal of Consciousness Studies*, 21: 6-32.

Hobson, J.A. și McCarley, R. (1977), The brain as a dream state generator: an activation-synthesis hypothesis of the dream process. *American Journal of Psychiatry*, 134: 1335-1348.

Hobson, J.A., McCarley, R. și Wyzinski, P. (1975), Sleep cycle oscillation: reciprocal discharge by two brainstem neuronal groups. *Science*, 189: 55-58.

Hochberg L., Bacher, D., Jarosiewicz, B. *et al.* (2012), Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. *Nature*, 485: 372-375.

Hohwy, J. (2012), Attention and conscious perception in the hypothesis testing brain. *Frontiers in Psychology*, 3: 96, doi.org/10.3389/fpsyg.2012.00096

Hohwy, J. (2013), *The Predictive Mind*. New York: Oxford University Press.

Holeckova, I., Fischer, C., Giard, M.-H. *et al.* (2006), Brain responses to subject's own name uttered by a familiar voice. *Brain Research*, 1082: 142-152.

Holstege, G., Georgiadis, J., Paans, A. *et al.* (2003), Brain activation during human male ejaculation. *Journal of Neuroscience*, 23: 9185-9193.

Horikawa, T., Tamaki, M., Miyawaki, Y. și Kamitani, Y. (2013), Neural decoding of visual imagery during sleep. *Science*, 340: 639-642.

Hsieh, P.-J. și Tse, P. (2006), Illusory color mixing upon perceptual fading and filling-in does not result in „forbidden colors". *Vision Research*, 46: 2251-2258.

Hume, D. (1748), *Philosophical Essays Concerning Human*

Understanding. Londra: A. Millar.
Hurley, M., Dennett, D. și Adams, R. (2011), *Inside Jokes: Using Humor to Reverse-Engineer the Mind.* Cambridge, MA: MIT Press.
Ingvar, D. (1985), „Memory of the future": an essay on the temporal organization of conscious awareness. *Human Neurobiology*, 4: 127-136.
Jackson, F. (1982), Epiphenomenal qualia. *Philosophical Quarterly*, 32: 127-136.
Jackson, F. (1995), Postscript on „What Mary Didn't Know". În P. Moser și J. Trout (eds.), *Contemporary Materialism*. Londra: Routledge, pp. 184-189.
Jaspers, K. (1963), *General Psychopathology*. Chicago: University of Chicago Press.
Jaynes, E. (1957), Information theory and statistical mechanics. *Physical Review*, 106: 620-630.
Jouvet, M. (1965), Paradoxical sleep: a study of its nature and mechanisms. *Progress in Brain Research*, 18: 20-62.
Joyce, J. (2008), Bayes' theorem. *Stanford Encyclopedia of Philosophy*.
Julesz, B. (1971), *Foundations of Cyclopean Perception*. Chicago: University of Chicago Press.
Kandel, E. (1998), A new intellectual framework for psychiatry. *American Journal of Psychiatry*, 155: 457-469.
Kandel, E. (1999), Biology and the future of psychoanalysis: a new intellectual framework for psychiatry revisited. *American Journal of Psychiatry*, 156: 505-524.
Kant, I. (1790), Kritik der Urteilskraft. *Kants gesammelte Schriften*, 5. Berlin: Walter de Gruyter.
Kaplan-Solms, K. și Solms, M. (2000), *Clinical Studies in Neuro-*

Psychoanalysis: Introduction to a Depth Neuropsychology. Londra: Karnac.

Kihlstrom, J. (1996), Perception without awareness of what is perceived, learning without awareness of what is learned. În M. Velmans (ed.), *The Science of Consciousness: Psychological, Neuropsychological and Clinical Reviews*. Londra: Routledge, pp. 23-46.

Knill, J. și Pouget, A. (2004), The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences*, 27: 712-719.

Koch, C. (2004), *The Quest for Consciousness: A Neurobiological Approach*. Englewood, CO: Roberts and Company.

Kopelman, M., Bajo, A. și Fotopoulou, A. (2015), Confabulation: memory deficits and neuroscientific aspects. În J. Wright (ed.), *International Encyclopedia of Social and Behavioral Sciences*. New York: Elsevier.

Krahe, C., Springer, A., Weinman, J. și Fotopoulou, A. (2013), The social modulation of pain: others as predictive signals of salience – a systematic review. *Frontiers in Human Neuroscience*, 7: 386.

Kurzweil, R. (2005), *The Singularity is Near: When Humans Transcend Biology*. New York: Viking.

Lane, N. (2015), *The Vital Question: Why is Life the Way It Is?* Londra: Profile.

Lavie, P., Pratt, H., Scharf, B., Peled, R. și Brown, J. (1984), Localized pontine lesion: nearly total absence of REM sleep. *Neurology*, 34: 118-120.

LeDoux, J. (1996), *The Emotional Brain*. New York: Simon and Schuster.

LeDoux, J. (1999), Psychoanalytic theory: clues from the brain.

Neuropsychoanalysis, 1: 44-49.
LeDoux, J. și Brown, R. (2017), A higher-order theory of emotional consciousness. *Proceedings of the National Academy of Science*, 114: e2016-e2025.
Lee, J., Everitt, B. și Thomas, K. (2004), Independent cellular processes for hippocampal memory consolidation and reconsolidation. *Science*, 304: 839-843.
Léna, I., Parrot, S., Deschaux, O. *et al.* (2005), Variations in extracellular levels of dopamine, noradrenaline, glutamate, and aspartate across the sleep-wake cycle in the medial prefrontal cortex and nucleus accumbens of freely moving rats. *Journal of Neuroscience Research*, 81: 891-899.
Leng, G. (2018), *The Heart of the Brain: The Hypothalamus and Its Hormones*. Cambridge, MA: MIT Press.
LeVay, S. (1993), *The Sexual Brain*. Cambridge, MA: MIT Press.
Levine, J. (1983), Materialism and qualia: the explanatory gap. *Pacific Philosophical Quarterly*, 64: 354-361.
Libet, B., Gleason, C., Wright E. și Pearl, D. (1983), Time of conscious intention to act in relation to onset of cerebral activity (readinesspotential): the unconscious initiation of a freely voluntary act. *Brain*, 106: 623-642.
Lichtheim, L. (1885), On aphasia. *Brain*, 7: 433-484.
Liepmann, H. (1900), Das Krankheitsbild der Apraxie („motorischen Asymbolie") auf Grund eines Falles von einseitiger Apraxie. *Monatsschrift für Psychiatrie und Neurologie*, 8: 15-44.
Lightman, A. (2018), *Searching for Stars on an Island in Maine.* New York: Pantheon.
Lin, P., Abney, K. și Bekey, G. (2011), *Robot Ethics*. Cambridge, MA: MIT Press.

Linnman, C., Moulton, E., Barmettler, G. *et al.* (2012), Neuroimaging of the periaqueductal gray: state of the field. *Neuroimage*, 60: 505-522.
Lisman, J. și Buzsaki, G. (2008), A neural coding scheme formed by the combined function of gamma and theta oscillations. *Schizophrenia Bulletin*, 34: 974-980.
Lissauer, H. (1890), Ein Fall von Seelenblindheit, nebst einem Beitrag zur Theorie derselben. *Archiv für Psychiatrie und Nervenkrankheiten*, 21: 222-270.
Lupyan, G. (2015), Cognitive penetrability of perception in the age of prediction: Predictive systems are penetrable systems. *Review of Philosophy and Psychology*, 6: 547-569.
Lupyan, G. și Clark, A. (2015), Words and the world: predictive coding and the language-perception-cognition interface. *Current Directions in Psychological Science*, 24: 279-284.
Lupyan, G. și Thompson-Schill, S. (2012), The evocative power of words: activation of concepts by verbal and nonverbal means. *Journal of Experimental Psychology – General*, 141: 170-186.
Lupyan, G. și Ward, E. (2013), Language can boost otherwise unseen objects into visual awareness. *Proceedings of the National Academy of Sciences*, 110: 14196-14201.
Malcolm-Smith, S., Koopowitz, S., Pantelis, E. și Solms, M. (2012), Approach/avoidance in dreams. *Cognition and Consciousness*, 21: 408-412.
Man, K. și Damasio, A. (2019), Homeostasis and soft robotics in the design of feeling machines. *Nature Machine Intelligence*, 1: 446-452, doi.org/10.1038/s42256–019–0103–7
Maturana, H. și Verela, F. (1972), *Autopoiesis and Cognition: The Realization of the Living*. Londra: Dordrecht.

Marc, J., și Llinás, R. (1994), Human oscillatory brain activity near 40 Hz coexists with cognitive temporal binding. *Proceedings of the National Academy of Sciences*, 91: 11748-11751.

Mathur, P., Lau B. și Guo, S. (2011), Conditioned place preference behavior in zebrafish. *Nature Protocols*, 6: 338-345.

Maxwell, J. (1872), *Theory of Heat*. Londra: Longmans, Green and Co.

Mazur, J.E. (2013), Basic principles of operant conditioning. *Learning and Behavior*, 7th edn. New York: Pearson, pp. 101-126.

McCarley, R. și Hobson, J.A. (1977), The neurobiological origins of psychoanalytic dream theory. *American Journal of Psychiatry*, 134: 1211-1221.

McKeever, W. (1986), Tachistoscopic methods in neuropsychology. În H.J. Hannay (ed.), *Experimental Techniques in Human Neuropsychology*. Oxford: Oxford University Press, pp. 167-211.

Merker, B. (2007), Consciousness without cerebral cortex: a challenge for neuroscience and medicine. *Behavioral and Brain Sciences*, 30: 63-68.

Mesulam, M.M. (2000), Behavioral neuroanatomy: large-scale networks, association cortex, frontal syndromes, the limbic system, and hemispheric specializations. În *Principles of Behavioral and Cognitive Neurology*, 2nd edn. New York: Oxford University Press, pp. 1-120.

Meynert, T. (1867), Der Bau der Gross-Hirnrinde und seine örtliche Verschiedenheiten, nebst einem pathologisch-anatomischen Corollarium. *Vierteljahrsschrift für Psychiatrie*

in ihren Beziehungen zur Morphologie und Pathologie des Central-Nervensystems, die physiologischen Psychologie, Statistik und gerichtlichen Medizin, 1: 77-93, 119-124.

Meynert, T. (1884), *Psychiatrie: Klinik der Erkrankungen des Vorderhirns*. Viena: W. Braumuller.

Misanin, J., Miller, R. și Lewis, D. (1968), Retrograde amnesia produced by electroconvulsive shock after reactivation of a consolidated memory trace. *Science*, 160: 203-204.

Mohr, M. von and Fotopoulou, A. (2017), The cutaneous borders of interoception: active and social inference of pain and pleasure on the skin. În M. Tsakiris și H. de Preester (eds.), *The Interoceptive Basis of the Mind*. Oxford: Oxford University Press.

Mongillo, G. Barak, O. și Tsodyks, M. (2008), Synaptic theory of working memory. *Science*, 319: 1543-1546.

Montague, P., Dolan R., Friston, K. și Dayan P. (2012), Computational psychiatry. *Trends in Cognitive Sciences*, 16: 72-80.

Moruzzi, G. și Magoun, H. (1949), Brain stem reticular formation and activation of the EEG. *Electroencephalography and Clinical Neurophysiology*, 1: 455-473.

Motta, S., Carobrez, A. și Canteras, N. (2017), The periaqueductal gray and primal emotional processing critical to influence complex defensive responses, fear learning and reward seeking. *Neuroscience and Biobehavioral Reviews*, 76: 39-47.

Moustafa, A., Sherman, S. și Frank, M. (2008), A dopaminergic basis for working memory, learning and attentional shifting in Parkinsonism. *Neuropsychologia*, 46: 3144-3156.

Mulert, C., Menzinger, E., Leicht, G. *et al.* (2005), Evidence for a close relationship between conscious effort and anterior

cingulate cortex activity. *International Journal of Psychophysiology*, 56: 65-80.
Munk, H. (1878), Weiteres zur Physiologie des Sehsphäre der Grosshirnrinde. *Deutsche medizinische Wochenschrift*, 4: 533-536.
Munk, H. (1881), *Über die Functionen der Grosshirnrinde: gesammelte Mittheilungen aus den Jahren 1877-80*. Berlin: Albrecht Hirschwald.
Nader, K., Schafe, G. și LeDoux, J. (2000), Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. *Nature*, 406: 722-726.
Nagel, T. (1974), What is it like to be a bat? *Philosophical Review*, 83: 435-450.
Newman, J. și Baars, B. (1993), A neural attentional model for access to consciousness: a global workspace perspective. *Concepts in Neuroscience*, 4: 255-290.
Niedenthal, P. (2007), Embodying emotion. *Science*, 316: 1002-1005.
Nishimoto, S., Vu, A., Naselaris, T. *et al.* (2011), Reconstructing visual experiences from brain activity evoked by natural movies. *Current Biology*, 21: 1641-1646.
Nour, M. și Carhart-Harris, R. (2017), Psychedelics and the science of self-experience. *British Journal of Psychiatry*, 210: 177-179.
Nummenmaa, L., Hari, R., Hietanen, J. și Glerean, E. (2018), Maps of subjective feelings. *Proceedings of the National Academy of Sciences*, 115: 9198-9203.
Oakley, D. și Halligan, P. (2017), Chasing the rainbow: the non-conscious nature of being. *Frontiers in Psychology*, 8: 1924, doi.org/ 10.3389/fpsyg.2017.01924

Oberauer, K., Souza, A., Druey, M. și Gade, M. (2013), Analogous mechanisms of selection and updating in declarative and procedural working memory: experiments and a computational model. *Cognitive Psychology*, 66: 157-211.

Okuda J., Fujii, T., Ohtake, H., Tsukiura, T., Tanji, K., Suzuki, K., Kawashima, R., Fukuda, H., Itoh, M. și Yamadori, A. (2003), Thinking of the future and past: the roles of the frontal pole and the medial temporal lobes. *Neuroimage*, 19: 1369-1380.

Pace-Schott, E. și Hobson, J.A. (1998), Review of „The Neuropsychology of Dreams". *Trends in Cognitive Sciences*, 2: 199-200.

Paloyelis, Y., Krahe, C., Maltezos, S., Williams, S., Howard, M. și Fotopoulou, A. (2016), The analgesic effect of oxytocin in humans: a double-blind, placebo-controlled cross-over study using laser-evoked potentials. *Journal of Neuroendocrinology*, 28: 10.1111/jne.12347

Panksepp, J. (1974), Hypothalamic regulation of energy balance and feeding behavior. *Federation Proceedings*, 33: 1150-1165.

Panksepp, J. (1998), *Affective Neuroscience: The Foundations of Human and Animal Emotions.* New York: Oxford University Press.

Panksepp, J. (2011), The basic emotional circuits of mammalian brains: do animals have affective lives? *Neuroscience and Biobehavioral Reviews*, 35: 1791-1804.

Panksepp, J. și Biven, L. (2012), *The Archaeology of Mind: Neuroevolutionary Origins of Human Emotions*. New York: Norton.

Panksepp, J. și Burgdorf, J. (2003), „Laughing" rats and the evolutionary antecedents of human joy. *Physiology and Behavior*, 79: 533-547.

Panksepp, J. și Solms, M. (2012), What is neuropsychoanalysis? Clinically relevant studies of the minded brain. *Trends in Cognitive Sciences*, 16: 6-8.

Parr, T. și Friston, K. (2018), The anatomy of inference: generative models and brain structure. *Frontiers in Computational Neuroscience*, 12: 90, doi.org/10.3389/fncom.2018.00090

Partridge, M. (1950), *Pre-Frontal Leucotomy: A Survey of 300 Cases Personally Followed for 1-3 Years*. Oxford: Blackwell.

Parvizi, J. și Damasio, A. (2003), Neuroanatomical correlates of brainstem coma. *Brain*, 126: 1524-1536.

Pasley, B., David, S., Mesgarani, N. *et al.* (2012), Reconstructing speech from human auditory cortex. *PLoS Biology*, 10(1): e1001251, doi.org/10.1371/journal.pbi0.1001251

Pellis, S. și Pellis, V. (2009), *The Playful Brain: Venturing to the Limits of Neuroscience*. Oxford: One World.

Penfield, W. și Jasper, H. (1954), *Epilepsy and the Functional Anatomy of the Human Brain*. Little, Brown and Co.

Petkova, V. și Ehrsson, H. (2008), If I were you: perceptual illusion of body swapping. *PLoS One*, 3: e3832, doi.org/10.1371/journal.pone.0003832

Pezzulo, G. (2014), Why do you fear the bogeyman? An embodied predictive coding model of perceptual inference. *Cognitive Affective and Behavioral Neuroscience*, 14: 902-911.

Pfaff, D. (2005), *Brain Arousal and Information Theory*. Cambridge, MA, Harvard University Press.

Picard, F. și Friston, K. (2014), Predictions, perception, and a sense of self. *Neurology*, 83: 1112-1118.

Popper, K. (1963), *Conjectures and Refutations*. Londra: Routledge.

Qin, P., Di, H., Liu, Y. *et al.* (2010), Anterior cingulate activity and the self in disorders of consciousness. *Human Brain Mapping*, 31: 1993-2002.
Ramachandran, V.S. (1992), Filling in the blind spot. *Nature*, 356: 115.
Ramachandran, V.S. și Gregory, R. (1991), Perceptual filling in of artificially induced scotomas in human vision. *Nature*, 350: 699-702.
Ramachandran, V.S., Gregory, R. și Aiken, W. (1993), Perceptual fading of visual texture borders. *Vision Research*, 33: 717-721.
Reggia, J. (2013), The rise of machine consciousness: studying consciousness with computational models. *Neural Networks*, 44: 112-131.
Riggs, L. și Ratliff, F. (1951), Visual acuity and the normal tremor of the eyes. *Science*, 114: 17-18.
Roepstorff, A. și Frith, C. (2004), What's at the top in the top-down control of action? Script-sharing and „top-top" control of action in cognitive experiments. *Psychological Research*, 68: 189-198.
Rolls, E. (2014), *Emotion and Decision-Making Explained*. New York: Oxford University Press.
Rolls, E. (2019), Emotion and reasoning in human decision-making. *Economics*, 13: 1-31.
Rosenthal, D. (2005), *Consciousness and Mind*. Oxford: Oxford University Press.
Rovelli, C. (2014), *Seven Brief Lessons in Physics*. Londra: Allen Lane.
Runes, D. (1972), *Dictionary of Philosophy*. Totowa, NJ: Littlefield, Adams and Co.

Sacks, O. (1970), *Migraine*. Londra: Vintage.
Sacks, O. (1973), *Awakenings*. Londra: Duckworth [ed.rom.: *Revenirea la viaţă*, traducere din limba engleză de Mihaela Doagă. Bucureşti: Humanitas, 2012].
Sacks, O. (1984), *A Leg to Stand On*. New York: Simon and Schuster [ed.rom.: *Un picior de sprijin*, traducere din engleză şi note de Ioana Ilie. Bucureşti: Humanitas, 2013].
Sacks, O. (1985), *The Man Who Mistook His Wife for a Hat*. Londra: Duckworth [ed.rom.: *Omul care îşi confunda soţia cu o pălărie*, traducere din engleză de Dan Rădulescu. Bucureşti: Humanitas, 2005].
Schacter, D., Addis, D. și Buckner, R. (2007), The prospective brain: remembering the past to imagine the future. *Nature Reviews Neuroscience*, 8: 657-661.
Schindler, R. (1953), Das Traumleben der Leukotomierten. *Wiener Zeitschrift für Nervenheilkunde*, 6: 330.
Searle, J. (1980), Minds, brains, and programs. *Behavioral and Brain Sciences*, 3: 417-424.
Searle, J. (1992), *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.
Searle, J. (1993), The problem of consciousness. *Social Research*, 60: 3-16.
Searle, J. (1997), *The Mystery of Consciousness*. Londra: Granta.
Seeley, W., Menon, V., Schatzberg, A. *et al.* (2007), Dissociable intrinsic connectivity networks for salience processing and executive control. *Journal of Neuroscience*, 27: 2349-2356.
Seth, A. (2013), Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, 17: 565-573.
Shallice, T. (1988), *From Neuropsychology to Mental Structure*. Cambridge: Cambridge University Press.

Shannon, C. (1948), A mathematical theory of communication. *Bell System Technical Journal*, 27: 379-423.

Shapiro, L. (2007), *Princess Elisabeth of Bohemia and Rene Descartes: The Correspondence Between Princess Elisabeth of Bohemia and Rene Descartes.* Chicago: University of Chicago Press.

Sharf, B., Moskovitz, C., Lupton, M. și Klawans, H. (1978), Dream phenomena induced by chronic levodopa therapy. *Journal of Neural Transmission*, 43: 143-151.

Shewmon, D., Holmes, G. și Byrne, P. (1999), Consciousness in congenitally decorticate children: developmental vegetative state as self-fulfilling prophecy. *Developmental Medicine and Child Neurology*, 41: 364-374.

Skinner, B.F. (1953), *Science and Human Behavior*. New York: Macmillan.

Solms, M. (1991), Summary and discussion of the paper „The neuropsychological organisation of dreaming: implications for psychoanalysis". *Bulletin of the Anna Freud Centre*, 16: 149-165.

Solms, M. (1995), New findings on the neurological organization of dreaming: implications for psychoanalysis. *Psychoanalytic Quarterly*, 64: 43-67.

Solms, M. (1996), Was sind Affekte? *Psyche*, 50: 485-522.

Solms, M. (1997a), *The Neuropsychology of Dreams: A Clinico-Anatomical Study*. Mahwah, NJ: Lawrence Erlbaum Associates.

Solms, M. (1997b), What is consciousness? *Journal of the American Psychoanalytic Association*, 45: 681-778.

Solms, M. (1998), Before and after Freud's „Project". În R. Bilder și F. LeFever (eds.), Neuroscience of the Mind on the

Centennial of Freud's Project for a Scientific Psychology. *Annals of the New York Academy of Sciences*, 843: 1-10.

Solms, M. (2000a), Dreaming and REM sleep are controlled by different brain mechanisms. *Behavioral and Brain Sciences*, 23: 843-850.

Solms, M. (2000b), Freud, Luria and the clinical method. *Psychoanalysis and History*, 2: 76-109.

Solms, M. (2000c), A psychoanalytic perspective on confabulation. *Neuropsychoanalysis*, 2: 133-138.

Solms, M. (2001), The neurochemistry of dreaming: cholinergic and dopaminergic hypotheses. În E. Perry, H. Ashton și A. Young (eds.), *The Neurochemistry of Consciousness*. New York: John Benjamins, pp. 123-131.

Solms, M. (2008), What is the „mind"? A neuro-psychoanalytical approach. În D. Dietrich, G. Fodor, G. Zucker și D. Bruckner (eds.), *Simulating the Mind: A Technical Neuropsychoanalytical Approach*. Viena: Springer Verlag, pp. 115-122.

Solms, M. (2011), Neurobiology and the neurological basis of dreaming. În P. Montagna și S. Chokroverty (eds.), *Handbook of Clinical Neurology*, 98 (3rd series), *Sleep Disorders*, Part 1. New York: Elsevier, pp. 519-544.

Solms, M. (2013), The conscious id. *Neuropsychoanalysis*, 14: 5-85.

Solms, M. (2015a), *The Feeling Brain: Selected Papers on Neuropsychoanalysis*. Londra: Karnac.

Solms, M. (2015b), Reconsolidation: turning consciousness into memory. *Behavioral and Brain Sciences*, 38: 40-41.

Solms, M. (2017a), Empathy and other minds – a neuropsychoanalytic perspective and a clinical vignette. În

V. Lux și S. Weigl (eds.), *Empathy: Epistemic Problems and Cultural-Historical Perspectives of a Cross-Disciplinary Concept*. Londra: Palgrave Macmillan, pp. 93-114.

Solms, M. (2017b), Consciousness by surprise: a neuropsychoanalytic approach to the hard problem. În R. Poznanski, J. Tuszynski și T. Feinberg (eds.), *Biophysics of Consciousness: A Foundational Approach*. New York: World Scientific, pp. 129-148.

Solms, M. (2017c), What is „the unconscious", and where is it located in the brain? A neuropsychoanalytic perspective. *Annals of the New York Academy of Sciences*, 1406: 90-97.

Solms, M. (2018a), Review of A. Damasio, „The Strange Order of Things". *Journal of the American Psychoanalytic Association*, 66: 579-586.

Solms, M. (2018b), The scientific standing of psychoanalysis. *British Journal of Psychiatry – International*, 15: 5-8.

Solms, M. (2019a), The hard problem of consciousness and the Free Energy Principle. *Frontiers in Psychology*, 10: 2714, doi.org/10.3389/fpsyg.2018.02714

Solms, M. (2019b), Commentary on Edmund Rolls: „Emotion and reason in human decision-making". *Economics Discussion Papers*, No. 2019-45, Kiel: Institute for the World Economy.

Solms, M. (2020a), Dreams and the hard problem of consciousness. În S. Della Salla (ed.), *Encyclopedia of Behavioral Neuroscience*. New York: Oxford University Press.

Solms, M. (2020b), New project for a scientific psychology: general scheme. *Neuropsychoanalysis*, 21.

Solms, M. (în tipar), Notes on some technical terms whose translation calls for comment. În M. Solms (ed.), *Revised*

Standard Edition of the Complete Psychological Works of Sigmund Freud, 24. Lanham, MD: Rowman and Littlefield.

Solms, M. și Friston, K. (2018), How and why consciousness arises: some considerations from physics and physiology. *Journal of Consciousness Studies*, 25: 202-238.

Solms, M., Kaplan-Solms, K. și Brown, J.W. (1996), Wilbrand's case of „mind-blindness". În C. Code, C.-W. Walesch, A.-R. Lecours și Y. Joanette (eds.), *Classic Cases in Neuropsychology*. Hove: Erlbaum, pp. 89-110.

Solms, M., Kaplan-Solms, K., Saling, M. și Miller, P. (1988), Inverted vision after frontal lobe disease. *Cortex*, 24: 499-509.

Solms, M. și Panksepp, J. (2010), Why depression feels bad. În E. Perry, D. Collerton, F. LeBeau și H. Ashton (eds.), *New Horizons in the Neuroscience of Consciousness*. Amsterdam: John Benjamins, pp. 169-179.

Solms, M. și Saling, M. (1986), On psychoanalysis and neuroscience: Freud's attitude to the localizationist tradition. *International Journal of Psychoanalysis*, 67: 397-416.

Solms, M. și Saling, M. (1990), *A Moment of Transition: Two Neuroscientific Articles by Sigmund Freud*. Londra: Karnac.

Solms, M. și Turnbull, O. (2002), *The Brain and the Inner World: An Introduction to the Neuroscience of Subjective Experience.* Londra: Karnac.

Solms, M. și Turnbull, O. (2011), What is neuropsychoanalysis? *Neuropsychoanalysis*, 13: 133-145.

Solms, M. și Zellner, M. (2012), Freudian drive theory today. În A. Fotopoulou, D. Pfaff și M. Conway (eds.), *From the Couch to the Lab: Trends in Psychodynamic Neuroscience*. New

York: Oxford University Press, pp. 49-63.
Squire, L. (2009), The legacy of Patient HM for neuroscience. *Neuron*, 61: 6-9.
Stein, T. și Sterzer, P. (2012), Not just another face in the crowd: detecting emotional schematic faces during continuous flash suppression. *Emotion*, 12: 988-996.
Stoerig, P. și Barth, E. (2001), Low level phenomenal vision despite unilateral destruction of primary visual cortex. *Consciousness and Cognition*, 10: 574-587.
Strawson, G. (2006), Realistic monism – why physicalism entails panpsychism. *Journal of Consciousness Studies*, 13: 3-31.
Sulloway, F. (1979), *Freud: Biologist of the Mind*. New York: Burnett.
Szpunar, K. (2010), Episodic future thought: an emerging concept. *Perspectives on Psychological Science*, 5: 142-162.
Szpunar, K., Watson, J. și McDermott, K. (2007), Neural substrates of envisioning the future. *Proceedings of the National Academy of Sciences*, 104: 642-647.
The Public Voice Coalition (2018), Universal guidelines for Artificial Intelligence, Draft 9, 23 October. Bruxelles: Electronic Privacy Information Center; https://thepublicvoice.org/ai-universal-guidelines
Thorndike, E. (1911), *Animal Intelligence*. New York: Macmillan.
Tononi, G. (2012), Integrated information theory of consciousness: an updated account. *Archives of Italian Biology*, 150: 56-90.
Tossani, E. (2013), The concept of mental pain. *Psychotherapy and Psychosomatics*, 82: 67-73.
Tozzi, A., Zare, M. și Benasich, A. (2016), New perspectives on spontaneous brain activity: dynamic networks and energy

matter. *Frontiers in Human Neuroscience*, 10: 247, doi.org/10.3389/fnhum. 2016.00247

Tranel, D., Gullickson, G., Koch, M. și Adolphs, R. (2006), Altered experience of emotion following bilateral amygdala damage. *Cognitive Neuropsychiatry*, 11: 219-232.

Turnbull, O. Berry, H. și Evans, C. (2004), A positive emotional bias in confabulatory false beliefs about place. *Brain and Cognition*, 55: 490-494.

Turnbull, O., Bowman, C., Shanker, S. și Davies, J. (2014), Emotion-based learning: insights from the Iowa Gambling Task. *Frontiers in Psychology*, 5: 162, doi.org/10.3389/fpsyg.2014.00162

Turnbull, O., Fotopoulou, A. și Solms, M. (2014), Anosognosia as motivated unawareness: the „defence" hypothesis revisited. *Cortex*, 61: 18-29.

Turnbull, O., Jenkins, S. și Rowley, M. (2004), The pleasantness of false beliefs: an emotion-based account of confabulation. *Neuropsychoanalysis*, 6: 5-16.

Turnbull, O. și Solms, M. (2007), Awareness, desire, and false beliefs. *Cortex*, 43: 1083-1090.

Uhlhaas, P. și Singer, W. (2010), Abnormal neural oscillations and synchrony in schizophrenia. *Nature Reviews Neuroscience*, 11: 100-113.

Van der Westhuizen, D., Moore, J., Solms, M., van Honk, J. (2017), Testosterone facilitates the sense of agency. *Consciousness and Cognition*, 56: 58-67.

Van der Westhuizen, D. și Solms, M. (2015), Social dominance in relation to the Affective Neuroscience Personality Scales. *Consciousness and Cognition*, 33: 90-111.

Varela, F., Thompson, E. și Rosch, E. (1991), *The Embodied Mind:*

Cognitive Science and Human Experience. Cambridge, MA: MIT Press.
Venkatraman, A., Edlow, B. și Immordino-Yang, M. (2017), The brainstem in emotion: a review. *Frontiers in Neuroanatomy*, 11: 15, doi.org/10.3389/fnana.2017.00015
Vertes, R. și Kocsis, B. (1997), Brainstem and diencephaloseptohippocampal systems controlling the theta rhythm of the hippocampus. *Neuroscience*, 81: 893-926.
Walker, M. (2017), *Why We Sleep*. Londra: Penguin.
Wang, X. și Krystal, J. (2014), Computational psychiatry. *Neuron*, 84: 638-654.
Weiskrantz, L. (2009), *Blindsight: A Case Study Spanning 35 Years and New Developments*. New York: Oxford University Press.
Weizenbaum, J. (1976), *Computer Power and Human Reason: From Judgment to Calculation.* New York: W.H. Freeman.
Wernicke, C. (1874), *Der aphasische Symptomencomplex. Eine psychologische Studie auf anatomischer Basis*. Breslau: M. Crohn and Weigert.
Wheeler, A. (1990), *A Journey into Gravity and Spacetime*. New York: W.H. Freeman.
White, B., Berg, D., Kan, J. *et al.* (2017), Superior colliculus neurons encode a visual saliency map during free viewing of natural dynamic video. *Nature Communications*, 8: 14263, doi.org/10.1038/ncomms14263
Whitty, C. și Lewin, W. (1957), Vivid daydreaming; an unusual form of confusion following anterior cingulectomy. *Brain*, 80: 72-76.
Wilbrand, H. (1887), *Die Seelenblindheit als Herderscheinung und ihre Beziehungen zur Homonymen Hemianopsie zur*

Alexie und Agraphie. Wiesbaden: J.F. Bergmann.

Wilbrand H. (1892), Ein Fall von Seelenblindheit und Hemianopsie mit Sectionsbefund. *Deutsche Zeitschrift für Nervenheilkunde*, 2: 361.

Yang, E. Zald, D. și Blake, R. (2007), Fearful expressions gain preferential access to awareness during continuous flash suppression. *Emotion*, 7: 882-886.

Yovell, Y., Bar, G., Mashiah, M. *et al.* (2016), Ultra-low-dose buprenorphine as a time-limited treatment for severe suicidal ideation: a randomized controlled trial. *American Journal of Psychiatry*, 173: 491-498.

Yu, C.K.-C. (2007), Cessation of dreaming and ventromesial frontal-region infarcts. *Neuropsychoanalysis*, 9: 83-90.

Zahavi, D. (2017), Brain, mind, world: predictive coding, neo-Kantianism, and transcendental idealism. *Husserl Studies*, 34: 47-61, doi.org/10.1007/s10743–017–9218-z

Zeki, S. (1993), *A Vision of the Brain*. Oxford: Blackwell.

Zellner, M., Watt, D., Solms, M. și Panksepp, J. (2011), Affective neuroscientific and neuropsychoanalytic approaches to two intractable psychiatric problems: why depression feels so bad and what addicts really want. *Neuroscience and Biobehavioral Reviews*, 35: 2000-2008.

Zeman, A. (2001), Consciousness. *Brain*, 124: 1263-1289.

Zhou, T., Zhu, H., Fan, Z. *et al.* (2017), History of winning remodels thalamo-PFC circuit to reinforce social dominance. *Science*, 357: 162-168.

Index*

A

a doua lege a termodinamicii *vezi* lege
Abbott, A. 336-337
Abney, K. 375-376
Absher, J. 330-331
Abu-Hassan, K. 374
acetilcolină *vezi* neuromodulatori/neurotransmițători
activare 31, 39, 107-108, 134-150, 152-155, 166-167, 200, 205-206, 209-212, 217-218, 232-234, 310-311, 313-317, 322-323, 333-334, 336-338, 341-342, 355-356, 365-366
    definiția termenului 314-315, 337-338
    vezi și afecte; precizie; stare de veghe
acuratețe *vezi* model generativ: acuratețe
Adams, R. 351-352, 367-368, 372-373
Addis, D. 367
adicție 54-55, 124-125, 296-297
adresare și răspuns la întrebări 166-169, 174, 176-177, 180-182, 194, 220, 345-346
afazie 74-75, 329-331
    vezi și Meynert
afect senzorial *vezi* afecte: senzoriale

afecte 13, 39, 56-57, 106-109, 144-147, 151-152, 154-155, 197, 246, 249, 271-272, 278, 291-292, 304-305, 309, 321-323, 336-337, 254-256, 372-373
ca variabile categoriale (calități) 201-204, 219-221, 291-292, 309-310, 332-333
clasificări 110-114
dezinhibate 64-65, 79-81
fixate/stabilizate 152-153, 229-231, 278-279, 348-349, 356-357, 363-364, *vezi și* cortex: stabilizează conștiința; energie latentă; memorie: pe termen scurt
senzoriale 105-106, 110-113, 117-118, 124-125, 151-152, 244-245, 367-369; *vezi și* dezgust; durere; surpriză
versus nevoi 107-110, 113-115, 147-148, 201, 211-214, 271-272, 278, 289-290
vezi și activare; emoții de bază; pulsiune; sentimente
Africa de Sud 17, 32-33, 48, 119-120, 205-206
agnozie 72-75, 329-330
vezi și Meynert
Aiken, W. 359-360
Ainley, V. 358-359
Alberini, C. 365-366
Alboni, P. 332-333
alegere 89-91, 108-110, 121-122, 128, 150-151, 154-155, 211-212, 251-252, 273, 309-310
vezi și valori; voluntar versus involuntar
alte minți *vezi* filosofie: problema altor minți
amânare 90-91, 202, 229-231, 278-279
amigdală *vezi* nuclei și alte structuri cerebrale
amnezie 48-49, 51-54, 75-76, 90-91, 150-151
vezi și confabulație; Meynert

amplificare postsinaptică *vezi* precizie
amplitudine *vezi* precizie
anesteziologie 68-69, 77, 132-134
animale nonumane 14, 111-113, 118-119, 243-244, 292, 294-299, 303-305, 328-329
    vezi și mamifere; oi; păsări; primate; vertebrate
animale supuse intervenției de decorticare cerebrală 63-65, 74-80, 84-85, 93-94, 148-150, 236-237, 340-341, 366-367
anosognozie 54-55
antropomorfism 111-112
apercepție *vezi* empirism
    definiția termenului 329
apraxie 74-75, 329-330
    vezi și Meynert
aria tegmentală ventrală *vezi* nuclei și alte structuri cerebrale
Aserinsky, E. 26-28, 321-322
Ashby, W.R. 170-171, 348-349
Astor, R. 319
atenție 113-114, 141-142, 148-149, 153-155, 213-214, 316, 355-356, 359-363, 365-368
Atkinson, R. 370-371
automatism 90-91, 108-110, 117-118, 121-122, 142-143, 147-148, 154-155, 189-190, 205-206, 229-231, 236-237, 263-264, 336-337
    ca starea ideală de funcționare cognitivă 122-123, 230-236, 310-311
autovalidare *vezi* sisteme cu autoorganizare
Axmacher, N. 319

B

Baars, B. 92-94, 255-256, 330-332
Bailey, P. 336-337
Bajo, A. 326-327
Bandler, R. 336-337
Baragwanath, spital, Soweto 48, 102
Barak, O. 370-371
Bargh, J. 90, 270-271, 331-332, 363
Barkay, P. 319
Barrett, L. Feldman 112-113, 122-123, 245, 333
Barth, E. 340-341
Bastos, A. 352-353
Bayes, T. 181-183
  regula 190-191, 194-196, 205-208, 228-229, 351-352, 354-355, 357-358
  teorema 182-184, 351-352
  vezi și probabilitate: posterioară/anterioară
Bechtel, W. 324-325
behaviorism 19-24, 26-29, 31-32, 96-99, 201-203, 293-294, 323-324, 364-365
  „cutia neagră” 20-22, 98-99
  recompensă și pedeapsă 21-22, 37- 38, 48-49, 53-54, 96-99, 293-294, 323-324, 364-365; *vezi și* eroare de predicție a recompensei
  teoria învățării 20-24, 96-99
  vezi și lege: a efectului
Bekey, G. 375-376
Benasich, A. 346, 352-353
Benson, D. 330-331
Bentley, B. 340-341
Berlin, H. 327-329

Berridge, K. 37-38, 323-324
Berry, H. 326-327
Besharati, S. 326-327
Bienenstock, E. 364-365
bit *vezi* informație: biți
Biven, L. 333-334, 337-338, 341-342, 359-360, 367-368
Blake, R. 361-362
Blake, Y. 333-334
Blakemore, S. 358-359
Blignaut, J. 361-362, 372-373
Block, N. 330-331
Blomstedt, P. 336-337
Bogen, J. 324-325, 331-332
Bohr, N. 345-346
Bouton, C. 374
Bowlby, J. 123-124, 334
Braun, A. 38-39, 69-70, 323-326
Breakspear, M. 359-360
Brentano, F. 348-349
briciul lui Occam *vezi* lege: (principiul) parcimoniei
Broca, P.P. 24-25, 329-330
Brooks, S. 319
Brown, H. 359
Brown, J. 326-327, 329-330
Brown, R. 330-331
Brucke, E. von 41-42, 307
Buckner, R. 367-368
Burgdorf, J. 335
Buzsaki, G. 359
Byrne, P. 327-328

C

calculatoare 21-23, 93-94, 137-138, 254-255, 267-268, 283-284, 289-290, 293-294, 298-299, 347-348, 375-376
calitatea de a fi incognoscibil 172-176, 180-181, 194, 215-216, 274, 372-373
vezi și zonă Markov
Cameron-Dow, C. 325-326
Campbell, A.W. 71-72, 328-329
canal versus stare, funcții 137-140, 203-204, 314-315, 336-337
Cannon, W. 314
Canteras, N. 338-339
cantitate versus calitate *vezi* filosofie: qualia
capacitate de răspuns 62-65, 67-69, 79-81, 134-135, 142-144, 313-315, 337-338
vezi și scala de comă Glasgow; stare de veghe vegetativă
Capogrosso, M. 281-283, 374
Carhart-Harris, R. 158-160, 300-301, 342-344, 359-360, 376
Carobrez, A. 338-339
Carvalho, G. 327-328, 339-340
cazuri clinice
22 ale lui Solms 83-85, 330-331
B. 57, 79-80, 105-106
domnul S. 47-55, 90-91, 96-97, 150-151, 153-154, 218-219, 239-240, 256-257, 273-274, 277, 340-341
fetița cu hidroanencefalie 62-68, 94-95
fiica unui coleg psihiatru 68-69
Fräulein G. 72-75
H.M. 90-91
pacientul lui Blomstedt 123-124, 132, 336-337
pacientul lui Claparède 90-91, 94-97, 121-122

pacientul lui Dollman 225-228
pacientul lui Sperry 90-92, 95-96
Phineas Gage 80-81
S.M. 119-120, 333-334
T.N. 65-67
unu al lui Whitty și Lewin 83-84
W. 80-83, 94-95
W.B. 225-226
CĂUTARE *vezi* emoții de bază
ceva cum e *vezi* filosofie
Chabris, C. 355-356
Chalmers, D. 67, 247-273, 278-283, 285-288, 308-309, 325-326, 331-332, 345-346, 368-374
principiul coerenței structurale 262- 265, 278-279, 281-282, 370-371
principiul dublului aspect 265-269, 278-279, 281-282, 346
principiul invarianței organizatorice 279-282
Changeux, J.P. 330-331, 359-360
Charcot, J.-M. 21-22, 329-330
Chartrand, T. 89-90, 270-271, 321-322, 363-364
Chew, Y. 340-341
chimie 178-179
vezi și neuromodulatori/neurotransmițători
cibernetică 169-171; *vezi și* feedback; semnale de eroare
ciclu percepție/acțiune 148-152, 154, 172-173, 194-195, 215-217
circuite
dopaminergice mezocorticale-mezolimbice 36-38, 53-54, 325-326; *vezi și* emoții de bază: CĂUTARE
nervul optic 48, 66-67, 147-148, 150-151, 224-225, 327-328; *vezi și* orbire: periferică

sistemul limbic (în general) 25-26, 192, 337-338, 367-368; *vezi și* nuclei și alte structuri cerebrale
Cisek, P. 359
civilizație 116-117, 243-244
Claparède, E. 90-91, 94-96, 121-122, 331-332
Clark, A. 238-239, 242-243, 340-341, 351-356, 360-361, 366-368
Coenen, A. 67, 327-328
Coenen, V. 333-334
cognitivism/revoluție cognitivă 21- 24, 26-27, 56-57, 76-77, 85, 96-97, 131, 167-168, 254-257, 325-326
cogniție *vezi și* gândire; limbaj; percepție
conștientă 70-71, 80-81, 83-84, 92-94, 96-97, 117-118, 142-143, 152-155, 212-213, 229-234, 238-239, 244-245, 248, 254-257, 270-272, 363-364, 365-366; *vezi și* afecte: fixate/stabilizate; memorie: pe termen scurt
inconștientă 14-15, 43, 76-77, 80- 81, 87-96, 121-122, 154-155, 229- 232, 235-236, 251-252, 270-272, 277, 288-289, 308-309, 328-329, 363-364, 365-366
versus afect 14-15, 51-52, 56-57, 76, 80-81, 94-96, 106-107, 121-123, 128-129, 142-144, 146-149, 151-155, 160, 197-198, 211-212, 223, 229-230, 235-236, 244-245, 249, 271-273, 278-280, 294-295
Colby, K. 375
Cole, S. 326-327
coliculi superiori *vezi* nuclei și alte structuri cerebrale
Collinger, J. 374
Coltheart, M. 326-327
comă 61, 63-65, 67-69, 74-76, 131, 134-135, 143-144, 160-162, 166-167, 285-286, 309, 372-373
complexitate *vezi* model generativ: complexitate

Conant, R. 348
confabulație 17-18, 48-55, 96-97, 150-151, 153-154, 256-258
conferințe despre știința conștiinței 47-38, 307, 325-326, 333-334
consolidare/reconsolidare 114-115, 140-141, 193-194, 228-236, 238-239, 252-254, 310-311, 363-366
  calitatea de a fi greu de uitat a memoriei nondeclarative 121-123, 193-194, 235-236
  vezi și automatism; context
conștiință *vezi și* cogniție: conștientă; percepție: conștientă; sentimente; stare de veghe
  accesare 78, 330-331; *vezi și* sine
  artificială 279-282, 285-303, 311-312, 375-376
  ca travaliu de predicție în curs 231-232, 238-239, 310-311, 360-362; *vezi și* cogniție: conștientă
  depinde de activarea trunchiului cerebral 131-137, 233-234, 300-301, 308-309, 324-325; *vezi și* activare
  este endogenă 57-58, 132-134, 137-138, 141-142, 151-152, 197-198, 217-218, 234-235, 304-305, 366-367, 371-372
  este fundamental afectivă 132-134, 151-152, 154-155, 197-198, 220-221, 271-272, 286-289, 296-297, 300-301, 308-309; *vezi și* afecte
  este incertitudine resimțită 212-213, 229-235, 236-237, 245, 291-292; *vezi și* incertitudine; sentimente
  este o formă de optimizare a preciziei 215-217, 229-230, 245, 294-295, 300-301, 356-359; *vezi și* precizie
  experiență (conținutul, calitatea) versus stare de veghe (nivelul, cantitatea) 26-27, 30-31, 39, 66-70, 76-78, 91-92, 128-129, 133-135, 308-309
  fenomenală (primară) versus reflexivă (secundară, gânduri de ordin superior) 78, 80-81, 330-331

fizică *vezi* fizică
origine 215-216, 218-221, 250, 252-253, 259-260, 262-266, 269-272, 274-275, 277, 286-288, 292, 308, 310-311
stări de ~ modificate 217-218
teorii 12, 92-94, 167-168, 255-259, 262-265, 278-279; *vezi și* eroarea privitoare la cortex
zori 13, 110-111, 220-221, 269-271, 278-279, 304-305
conștiință artificială *vezi* conștiință: artificială
context 108-109, 121-123, 145-147, 151-153, 189, 202-203, 290-292, 336-337, 341-342, 354-355
preconizat 149-154, 203-206, 209-212, 236-237, 239-240, 244-246, 309-310
vezi și percepție: contextualizează afectul
control respirator *vezi* funcții autonome/vegetative
Conway, M. 326-327
Cooper, L. 364-365
Coren, S. 365-366
Corlett, P. 359-360
cortex 13-15, 25-26, 29-35, 39, 56-59, 62-72, 74-78, 84-85, 87-89, 118-119, 131, 134-140, 143-144, 146-149, 151-153, 159-160, 165-166, 195-196, 223-234, 236-240, 243-245, 249-250, 260-261, 263-264, 273, 278, 281-283, 285-287, 294-295, 303-304, 308-309, 322-325, 328-333, 336-339, 341-344, 348-349, 352-353, 363-364
cingulat anterior 78, 81-85, 123-124, 334-336
frontal 25-26, 34-36, 39, 78-83, 92-95, 116-117, 225-227, 263-264, 282-283
funcționarea inconștientă 91-92, 136-138, 233-234, 273, 308-309, 324-325, 368-369
hipocamp 25-26, 90-91, 121-122, 239-240, 282-283, 336-337,

341-342, 365-368
insular 57, 78-80
limbaj 24-25, 33-34, 74-75, 77-78, 81-85, 138-140, 283-284, 327-328, 330-331
motor 33-34, 281-283
occipital 33-34, 65-67, 72-74, 78, 224-228, 234-235, 284-285, 359-360
parietal 34-36, 92-93, 146-147, 263-264
se specializează în incertitudine (complexitate, contexte) 230-231, 236-237, 366-367; *vezi și* voluntar versus involuntar
stabilizează conștiința 152-153, 229-231, 278-279, 363-364; *vezi și* amânare; memorie: pe termen scurt
temporal 74-75, 78, 139-140, 283-284; *vezi și* animale supuse intervenției de decorticare cerebrală; hidroanencefalie
cortex frontal/prefrontal *vezi* cortex
Craig, A.D. (Bud) 56-57, 79-80, 330-331
Crick, F. 93-94, 248-252, 271-272, 307-308, 321-322, 368-369
cristale 157-159, 170-172, 342-343
crize *vezi* epilepsie
Crucianelli, L. 358-359
Çukur, T. 367-368

D

Damasio, A. 14-15, 56-58, 64-65, 79-83, 105-106, 133-134, 142-143, 161-163, 299-300, 326-328, 330-331, 336-337, 339-340, 343-344, 372-373
Damasio, H. 56-57, 326-327, 330-331
Darwin, C. 112-113, 169-170, 347-348

Davies, P. 247, 358-359, 368-369
Davis, E. 336-337
Debiec, J. 365-366
Decety, J. 358-359
Deco, G. 359-360
Dehaene, S. 92-93, 330-332, 359-360
Dement, W. 27-28, 321-322
demonul lui Maxwell 358-359
Dennett, D. 42, 367-368, 372-373
Depaulis, A. 336-337
depresie 19, 54-55, 123-125, 132-134
Descartes, R. 56-57, 252-253, 264-265, 369-370
dezgust 110-113, 151-152, 203-204, 244-245
vezi și afecte: senzoriale
Ditchburn, R.W. 233-235, 365-366
Dolan, R. 359-360
Dollman, A. 227, 319, 372-373
Domhoff, W. 367-368
domnul S. *vezi* cazuri clinice
dopamină *vezi* circuite; neuromodulatori/neurotransmițători
DORINȚĂ SEXUALĂ *vezi* emoții de bază
Dudai, Y. 364-365
Duffy, E. 314
durere 13-14, 68-70, 95-97, 105-106, 110-111, 117-118, 132, 134-135, 145-147, 151-152, 232-323, 244-245, 272-273, 289-291, 354-356, 338-340
vezi și afecte: senzoriale

E

ecuații 177-178, 216-217, 242-243, 302-303, 342-343
acțiune 357-358
conștiință 357-358
divergență perceptivă 350-351
energie (în general) 342-343
energie liberă (Friston) 178-180, 215-216, 349, 357-369
energie liberă (Gibbs) 178-179
energie liberă (Helmholtz) 178-180
energie liberă preconizată 356-357
eroare de predicție 357-358
informație (entropie Shannon) 315-316
optimizarea percepției 216-217
percepție 357-358
surprindere 350-351
vezi și Bayes: teorema
Edelman, G. 92-93, 331-332
Edlow, B. 322-323, 337-339
EEG *vezi* electroencefalogramă
Ehrsson, H. 368-369
Einstein, A. 162-164, 203, 321-322, 343-344
Eisenberger, N. 335
Ekman, P. 333
electrocorticogramă 284-285
electroencefalogramă (EEG) 23-24, 32-33, 134-136, 165-167, 283-284, 313-314
desincronizată 134-135, 165-167
sincronizată 93-94, 134-135, 165-167, 258, 263-264, 341-342
undele alfa 134-136
undele beta 134-136, 352-353
undele delta 134-137

undele gama 93-94, 134-137, 258, 263-264, 352-353, 355-356
undele teta 134-136, 341-342; *vezi și* entropie: EEG
Elisabeta a Boemiei, prințesă 369-370
Ellis, G. 319, 372-373
emoție inconștientă *vezi* sentiment: versus „emoție inconștientă”
emoții de bază 111-113, 134-137, 141-142, 152-153, 243-244; *vezi și* afecte: clasificări; emoții secundare
CĂUTARE 37-38, 46-48, 53-54, 57-58, 115-117, 123-126, 132-134, 137-140, 145-148, 205, 230-231, 238-239, 310-311, 323-326, 333-334, 341-342, 345-346, 352-353, 355-356, 365-366
DORINȚĂ SEXUALĂ 64-65, 113-116, 124-125, 127, 144-146, 212-213, 244-245, 296-297, 333-334, 339-340
FRICĂ 98-99, 105-107, 113-114, 117-123, 132-134, 137-140, 145-146, 160-162, 199, 204-206, 231- 233, 235-236, 244, 290-291, 303- 305, 330-334, 338-340, 368-369
FURIE 64-65, 79-80, 116-119, 122-123, 127, 132-134, 137-138, 145-146, 160-161, 199, 244-245, 304-305, 333-334, 338-340
GRIJĂ 64-65, 79, 124-126, 141-142, 145-146, 303-305, 335, 339-340
JOACĂ 17-18, 62, 64-65, 79-80, 125-128, 243-244, 334-336
PANICĂ/MÂHNIRE 52-53, 98- 99, 105-107, 122-126, 132-134, 137-138, 145-147, 160-162, 203-204, 235-236, 244-246, 290-291, 303-304, 307, 334-336
emoții secundare 112-113, 122-123
empatie 127, 219-220, 243-244, 297-298, 335, 359-360
vezi și emoții de bază: JOACĂ
empirism 19-20, 61-62, 87, 96-97, 133-134, 244-245, 258, 292-293, 328-329
apercepție versus senzație brută 70-73, 75-76, 78, 87-88, 90-

91, 152-153, 194, 328-330
tabula rasa 70-73
vezi și Meynert; știință: reguli
encefalită
cu *herpes simplex* 79
letargică 24-25, 217-218
energie latentă 178-179, 289-290, 348-349, 356-357, 363-264
vezi și energie liberă; psihanaliză: procese secundare
energie liberă 12-13, 158-160, 183-185, 199-208, 212-220, 237-238, 241-242, 289-292, 309-310, 342-343, 348-352, 356-358, 363-364
definițiile termenului 178-180, 348-349
principiu 158-159, 162-163, 177-182, 184-185, 197, 200, 203-208, 212-213, 215-219, 230-231, 237- 238, 257, 288-289, 300-301, 342-344, 352-353, 366-367
vezi și ecuații
energie vitală 42, 45-46, 307, 324-325
England, J. 347-348
entropie 155, 162-172, 178-181, 187, 219-220, 269-270, 273-274, 286-289, 310-311, 315-316, 343-351
ca informație medie 165-166, 168-169, 178-180, 201-202, 286-287, 315-316, 349, 266-267
EEG 165-167
epilepsie 62, 77-78, 83-84, 90-91, 364-365
eroare de predicție a recompensei 231-233
eroarea privitoare la cortex 13-15, 57-89, 91-93, 131, 198-199, 223, 229-230, 273, 353-354
Ethier, C. 374
etică 13-14, 64-65, 68-69, 90-91, 111-112, 288-289, 396-397, 311-312, 376-377

eu *vezi* psihanaliză: ego; sine
Evans, C. 326-327
„Eve Periaqueduct”187-193, 195-196, 203-204, 208-211, 288-289
Everitt, B. 365-366
evoluție *vezi* selecție naturală
experiență estetică 245
    vezi și glume
explozie combinatorie 202-203, 210-211
    vezi și model generativ: complexitate
exterocepție 113-114, 146-147, 151-155, 197-198, 206-208, 212-214, 244-246, 309-311, 353-354, 357-358, 366-367, 371-372
    vezi și context; percepție; incertitudine: aplicată
Ezra, M. 337-338

F

feedback 147-150, 169-171, 173-174, 188, 190-191, 193-194, 197-198
    vezi și semnale de eroare
Feldman, H. 354-359
Ferrarelli, F. 358-359
Feynman, R. 374
filosofie 13, 21, 96-97, 107, 146, 155, 252-253, 259, 272, 302-303, 333
    argumentul cunoașterii 250, 272, 369-370, 373
    ceva cum e 25-26, 67-68, 77, 95, 200, 218-221, 248, 251, 256, 267-268, 273-274, 295, 308, 311; *vezi și* qualia
    dualism 218-219, 252, 264, 372-373
    epifenomenalism 251-252, 370; *vezi și* liber-arbitru
    inversarea qualia *vezi* filosofie: spectre inversate

metafizică 15, 19, 22-23, 25, 28-29, 46, 155, 253, 258
monism al aspectului dual 146, 155, 161-162, 200, 213, 253-254, 259-260, 262-263, 265-266, 280-281, 308, 311, 372
„nimeni acasă” 67, 285-286; *vezi și* ceva cum e; filosofie: zombi
panpsihism 94, 267-269, 272, 280-282
pe cine vizează intenția 152, 213, 337, 348
problema altor minți 20-21, 69-70, 77, 84-85, 174, 292-297, 348
problema fixării 92-94
problema minte-corp 19, 22-23, 25, 28-29, 44, 155, 201, 247-252, 259-260, 294, 308, 369-370; *vezi și* problema dificilă
qualia 11, 78, 201, 204, 219-221, 224, 244-246, 248, 251-252, 292, 348, 371, 373; *vezi și* conștiință; filosofie: ceva cum e; percepție: conștientă
reducționism 249, 252, 262-264, 307, 371
spectre inversate 107-108, 244, 251, 272, 373
teleologie 169-170, 176, 219-220, 347
vid epistemic/explicativ 250, 253, 257-260, 263, 272-273, 370
zombi 67, 69-70, 231, 233, 235, 272, 293-294, 354, 370
vezi și empirism; etică; funcționalism; problema dificilă; știință, reguli
Fischer, D. 131, 335-336
Fischmann, T. 325-326
FitzGerald, T. 358-359
fizică 15, 42, 45-46, 155, 160, 165, 167-168, 200, 262, 265-266, 279, 287-288, 332, 344-346
a conștiinței 162-163, 174, 198-221, 224, 247, 265-266, 269-270, 275, 287-288, 311
definiția termenului 342-343

vieții 13, 159, 162-163, 168-173, 269-270, 274-275, 288-289, 311-312, 347-348
vezi și entropie; materie; termodinamică; travaliu
Flechsig, P. 328-329
Fletcher, P. 359-360
Fliess, W. 45, 324-326, 321-322, 355-356, 364-365
Fonagy, P. 326-327
Forrester, G. 335-336
forță a semnalului *vezi* precizie
Fotopoulou, A. (Katerina) 300-301, 326-327, 356-359
Frank, J. 323-324
Frank, M. 358-359
Freeman, W. 35-36
Freud, S. 14-15, 20-21, 25-26, 28-32, 38-39, 41-48, 53-59, 63- 64, 72-73, 87-91, 95-98, 102, 106-108, 115-116, 145-146, 158-161, 184-185, 197-198, 214-215, 223, 229-232, 243-244, 270-271, 287-288, 307-311, 321-332, 337-338, 340-344, 348-349, 355-357, 361-365
ideea că „conștiința ia naștere acolo unde se sfârșește urma mnezică” 229-232, 310-311, 363-364
ideea de „cauză principală a mecanismului psihic” 44-45, 57, 88- 89
ideea de conștiință în „straturile intime” ale creierului 58-59, 145-146, 304-305
ideea despre prozencefal ca „ganglion simpatic” 44-45, 88-90, 197-198
Interpretarea viselor (1900) 28-29, 41-42
„Proiect de psihologie” (1895) 14- 15, 41-45, 159-160, 214-215, 307, 324-325, 355-357, 363-364
Zur Auffassung der Aphasien (1891) 44, 87-90, 329-331, 364-

365
vezi și psihanaliză
FRICĂ *vezi* emoții de bază
Friston, K. 12, 14-15, 157-163, 170-171, 174-185, 191-193, 214-216, 257-258, 273-274, 287-288, 300-301, 319, 340-343, 347-349, 351-360, 366-367, 371-372, 376-377
supa primordială simulată 157-159, 170-176, 180-182, 273-274
Frith, C. 242-243, 256-257, 367-368
funcții autonome/vegetative 61, 65-67, 107-109, 142-143
ciclul somn – stare de veghe 143-144, 313, 327-329, 336-337
controlul defecației 119-120
controlul respirator 104-110, 119-120, 144-147, 160-162, 168-170, 179-180, 195-198, 203-204, 304- 305, 332-333, 338-340
controlul urinării 108-109, 147-148, 205-206, 332-333
digestie 108-109
echilibru energetic 13, 98-99, 105-109, 160-162, 183-184, 200-203, 289-290, 352-353
echilibrul fluidelor 107-108, 146-147, 160-162, 201-203
presiune arterială 107-109, 145-146, 338-339
reflex vasovagal 104-106, 119-120, 122-124
ritm cardiac 119-120, 144-145, 332-333, 338-339
termoreglare 105-106, 108-109, 160-163, 179-180, 191, 197-198, 201, 220-221, 289-291, 294-295
vezi și somn
funcții vegetative *vezi* funcții autonome/vegetative
funcționalism 22-23, 43-44, 253-263, 273, 278-279, 308, 324-325, 370-371
vezi și conștiință: teorii
FURIE *vezi* emoții de bază

G

Galilei, Galileo 342-343
Galin, D. 331-332
Gallant, J. 284-285
Gallese, V. 56-57, 326-327
gama/teoria sincronizării sub 40 Hz *vezi* conștiință: teorii; *vezi și* electroencefalogramă
ganglioni bazali *vezi* nuclei și alte structuri cerebrale
Garcia-Rill, E. 336-337
Gates, W. (Bill) 375-376
gândire 117-118, 121-122, 128, 236-237, 239-244, 277, 284-286, 333-334, 374
vezi și hoinăreala minții
„generalizabilitate” *vezi* consolidare/reconsolidare: calitatea de a fi greu de uitat a memoriei nondeclarative; model generativ: „generalizabilitate”
Geschwind, N. 76-77, 341-342
Gilbert, J. 375-376
Ginsborg, B. 365-366
girusul cingular anterior *vezi* cortex
Gloor, P. 330-331
glume 245, 367-369
vezi și experiență estetică
Golaszewski, S. 336-337
Goltz, F. 75-76
Goodglass, H. 330-331
Gosseries, O. 345-346
Green, A. 326-327
Gregory, R. 353-354, 359-360
GRIJĂ *vezi* emoții de bază

Groddeck, G. 324-325
Groote Schuur, spital, Cape Town 17-18
Grubrich-Simitis, I. 326-327
Guo, S. 375-376

H

Haikonen, P. 375-376
Halligan, P. 369-370
Hallmann, E. 324-325
halucinație 28-29, 36-37, 74-75
    vezi și percepție: ca halucinație sau fantezie controlată
Harding, D. 340-341
Harlow, J. 330-331
Hartmann, E. 323-324
Hassin, R. 363-364
Havlik, M. 370-371, 376-377
Hawking, S. 343-344, 375-376
Hebb, D. 313-314, 364-366
Heever, D. van den 319, 372-373
Helmholtz, H. von 178-179, 194, 227-228, 258, 307, 311-312, 342-344, 348-349, 359-360, 365-366
hemoragie subarahnoidă 81-84, 322-323
Herff, C. 284-285, 373-374
Hering, E. 340-341
Hesselmann, G. 355-356
hidroanencefalie 61-70, 75-78, 84-85, 93-95, 148-150, 236-237, 294-297, 327-329, 340-341, 366-367
hipocamp *vezi* cortex
hipotalamus *vezi* nuclei și alte structuri cerebrale

Hobson, J.A. 29-32, 37-39, 46-48, 322-324, 335-337, 359-360, 366-367
Hochberg, L. 374-375
Hodson, R. 372-373
Hohwy, J. 150-151, 191-193, 340-341, 352-355, 358-363, 366-367
hoinăreala minții 236-239
vezi și mod implicit
Holeckova, I. 336-337
Holmes, G. 327-328
Holstege, G. 336-337
homeostază 57-58, 98-99, 110-111, 160-165, 168-171, 173-174, 176-177, 179-182, 184-185, 193-194, 196-198, 200-202, 230-231, 234-235, 245-246, 286-287, 289-291, 300-301, 304-305, 308-310, 313-314, 353-354
originea 175-176, 308-309
vezi și cibernetică; entropie; semnale de eroare; travaliu
Horáče, J. 370-371, 376-377
Horikawa, T. 374
hormoni *vezi* neuromodulatori/neurotransmițători
Horwood, T. 319
Hromnik, T. 372-373
Hsieh, P.-J. 365-366
Hume, D. 70-71, 84-85, 332-333
Humphrey, N. 326-327, 333-334
Hurley, M. 367-368, 372-373

I

ierarhie predictivă 154, 187-198, 205, 207-210, 220-221, 231-233, 237, 241, 243, 245-246, 288, 310, 344, 352-354, 358-359

a elementelor anterioare 190-192, 196
imagistică prin rezonanță magnetică funcțională (RMNf) 23-24, 132-133, 242-243, 284-285, 355-356
Immordino-Yang, M. 337-339
incertitudine 20, 99, 109-110, 112-113, 122, 150-154, 164-169, 178-181, 200, 202, 205-214, 229-236, 269-270, 275, 291-292, 295-296, 304-305, 310-311, 315-317, 351-352, 357-358, 362
aplicată 152-153, 212-213; *vezi și* percepție: conștientă
preconizată 152-154, 207-212, 355-356
resimțită *vezi* conștiință: este incertitudine resimțită
speri la ce e mai bine/te pregătești pentru ce e mai rău *vezi* sentimente: ghidează rezolvarea problemelor
inferență *vezi* predicție
informație 21-23, 44, 65-67, 76-78, 87-88, 136, 164-168, 178-181, 183-184, 233-234, 236-237, 255-257, 265-270, 281-284, 286-287, 314-317, 345-349, 371-373
biți (1/0, cifre binare) 137-140, 165-168, 346-347
byte 345
este fizică 22-23, 44, 166-168, 183-184, 265-267, 316-317, 346-348; *vezi și* materie
medie *vezi* entropie
reciprocă 180-181, 187, 352-253
informație reciprocă *vezi* informație
Ingvar, D. 328-329, 367-368
insecte 151-152, 303-304, 340-341
instinct (ghicire, intuiție) 65-67, 121-122, 217-218, 304-305, 333-334, 368-369
instinct și reflex 107-108, 112-126, 142-143, 149-150, 191-193, 199-200, 230-231, 234-235, 309-310, 313, 335-340, 375-376
vezi și automatism; funcții autonome/vegetative

inteligență 13-14
vezi și sentimente: versus inteligență
intenționalitate 43, 45-46, 52-53, 62-63, 84-85, 142-144, 159-160, 169-170, 181-182, 270-271, 292-293, 295-296, 309-310, 336-338, 372-373
origine 175-176, 184-185, 197, 219-221, 229-231, 288-289, 291-292
vezi și filosofie: pe cine vizează intenția
internet 93-94, 267-268, 272-273
interocepție 57, 76-80, 110-111, 206-208, 217-218, 357-358
introspecție *vezi* subiectivitate: rapoarte introspective
intuiție *vezi* instinct
ipseitate *vezi* subiectivitate
izvor ascuns 13-14, 99, 131, 304-305

Î

împărțire pe categorii/compartimentare a nevoilor *vezi* afecte: ca variabile categoriale
încredere *vezi* precizie
învățare din experiență 109-110, 112-116, 120-121, 125, 149-151, 153-154, 197, 200, 205-206, 209, 229- 232, 238-239, 309-310, 336-337; *vezi și* consolidare/reconsolidare; context; sentimente: ghidează re-zolvarea problemelor

J

Jackson, F. 369-370
Jackson, J. Hughlings 48
James, T. 319

Jasper, H. 77-78, 338-339
Jaspers, K. 330-331
Jaynes, E. Thompson 166-167, 345-346
Jenkins, S. 326-327
JOACĂ *vezi* emoții de bază
Journal of Consciousness Studies 258, 371-372
Jouvet, M. 29-31, 322
Joyce, J. 351-352
Julesz, B. 340-341

K

Kalaska, J. 359
Kandel, E. 55-56, 326-327
Kant, I. 169-170, 342-343
    vezi și calitatea de a fi incognoscibil
Kentridge, E. 319
Kentridge, S. 319
Kernberg, O. 327
Kihlstrom, J. 90, 270-271, 332
Kinsbourne, M. 327
Kleitman, E. 26-28, 322
Knill, J. 351-352
Koch, C. 93-94, 332-333
Kocsis, B. 341-342
Kopelman, M. 326-327
Kozakova, E. 371, 376
Krahe, C. 358-359
Krikler, J. 319
Kris, E. 55

Krystal, J. 359-360
Kurzweil, R. 474

L

Lake, E. 319
Lau, B. 375
Lavie, P. 322-323
LeDoux, J. 80-81, 121-122, 326-327, 330-331, 334-335, 364-365
Lee, J. 365-366
lege
    a afectului 97-98, 108-110, 200, 309-310, 376-377
    a efectului 20-22, 97-99
    a lui Friston 184-185, 196, 206-212, 216-217, 288-289, 310-311, 376-377
    a lui Hebb 233, 365, 370
    a lui Miller 253-254, 261
    a lui Ribot 253-254, 261
    a termodinamicii, prima 162-164
    a termodinamicii, a doua 162-164, 181-182, 316, 344-345, 349, 358-359
    (principiul) parcimoniei 203, 237-239, 354-355
legea lui Friston *vezi* lege
Léna, I. 323-324
Leng, G. 333
leucotomie/lobotomie 34-37, 39, 83-84
Leuzinger-Bohleber, M. 326-327
LeVay, S. 333
Levine, J. 257, 369-370
Lewin, W. 83-84, 330-331

Lewis, D. 364-365
liber-arbitru *vezi și* alegere; voluntar versus involuntar 247, 249, 251-252, 307, 351-352, 369-370
Libet, B. 252, 369-371
Lichtheim, L. 329-330
Liepmann, H. 329-330
Lightman, A. 344-345
limbaj 77, 84-85, 128, 240-244, 252-253, 256-257, 283-284, 328-331, 347-348, 366-368
vezi și cortex: limbaj; subiectivitate: rapoarte introspective
Lin, P. 375-376
Linnman, C. 337-338
Lisman, J. 358-359
Lissauer, H. 329-330
Llinás, R. 93-94, 326-327, 331-332
lobotomie *vezi* leucotomie
Locke, D. 70-71, 250-252, 272-273, 369-370, 374-375
Ludwig, K. 307
Lupyan, G. 240-242, 360-361, 367-368

M

Maclean, P. 76
Magoun, H. 131, 134, 336
Malcolm-Smith, S. 326
Malmo, R. 313-314
mamifere 13-14, 111-112, 118-119, 124-126, 292-295, 303-304, 328-329, 338-339
Marc, J. 332
Markov, A. 348

Markov, zonă 172-178, 180-181, 194-196, 215-217, 246, 269-270, 288-289, 349, 351-352
Martin, J. 319, 362, 373
„Mary” *vezi* filosofie: argumentul cunoașterii
matematică 42, 45-46, 159-160, 181-182, 184-185, 242-243, 253-254, 310-311, 313-317, 342-343
materie, nu drept concept fundamental 163-165, 342-348
Mathur, P. 375
Maturana, H. 348-349
Maxwell, J.C. 358
Mazur, J. 333
McCarley, R. 322
McKeever, W. 332
memorie
ca predicție 149-150, 152-153, 191-193, 230-233, 236-237, 239-241, 291-292, 303-304, 310-311
declarativă (explicită) 43, 121-122, 191-192, 231-237, 239-240, 285-286, 303-304, 310-311, 321-322, 365-367
nondeclarativă (implicită) 43, 120-122, 147-150, 191-192, 235-236, 270-271, 310-311, 321-322, 363-367; *vezi și* consolidare/reconsolidare: calitatea de a fi greu de uitat
pe termen scurt 23-24, 34-35, 80-81, 189-194, 229-237, 253-254, 291-292, 295-296, 310-311, 330-323, 363-364, 370-371
vezi și consolidare/reconsolidare; învățare din experiență
memorie de lucru *vezi* memorie: pe termen scurt
memorie declarativă *vezi* memorie
memorie emoțională *vezi* memorie: nondeclarativă; sentimente: versus la ce se referă
memorie episodică *vezi* memorie: declarativă
memorie nondeclarativă *vezi* memorie

memorie pe termen lung *vezi* memorie
memorie pe termen scurt *vezi* memorie
memorie procedurală *vezi* automatism: ca stare ideală de funcționare cognitivă; memorie: nondeclarativă
Merker, B. 14-15, 61-62, 64-65, 142-143, 146-147, 149-150, 152-153, 193-194, 220-221, 300-301, 327-328, 330-333, 338-342;
Mesulam, M.M. 336-337, 352-353
metafizică *vezi* filosofie
metodă experimentală *vezi* știință, reguli
Meynert, T. 71-72, 75-76, 84-85, 87, 134, 329-332
  asociere transcorticală 74-76, 92-93, 329-330
  cortical (voluntar, mintal) versus subcortical (reflex, fizic) 71-78, 87-89, 91-92, 329-330
  imagini mnezice 70-78, 87-89, 91-93, 149-152, 223, 235-237, 328-330; *vezi și* empirism
  înțelegere (cunoaștere) 70-76, 91-92, 250-251, 304-305
  mitologie referitoare la creier 84-85, 330-331
mezencefal *vezi* trunchi cerebral
Miller, P. 102-104
Miller, R. 364
Misanin, J. 364
Mishra, A. 373
mod implicit 116-117, 147-148, 154-155, 205, 230-231, 238-239, 310-311, 343-344, 352-353, 365-368
  vezi și emoții de bază: CĂUTAREA
model generativ 149-154, 174-175, 179-182, 187-188, 191-193, 205-206, 209-210, 213-214, 218-219, 225, 227, 230-231, 291-292, 353-355
  actualizare 180-183, 187-191, 193-194, 196, 205-209, 230-237, 310-311, 358-259, 363-365; *vezi și* Bayes, T.: regula

acurateţe 193-194, 203, 236-237, 349, 351-353, 366-367
complexitate 193-194, 201-203, 210-211, 236-238, 310-311, 349, 351-353, 366-367
eficienţă 236-239, 347-348, 352-353, 355-356
„generalizabilitate” 193-194, 203, 234-237, 310-311, 351-352
vezi şi context: preconizat; ierarhie predictivă; memorie; predicţie: în inferenţa activă
model predictiv *vezi* model generativ
Modell, A. 327
modulare *vezi* activare; neuromodulare versus neurotransmisie; precizie
Mohr, M. von 359
Mongillo, G. 370
monism al aspectului dual *vezi* filosofie
Montague, P. 360
mortalitate 19, 176-177, 247, 268-269, 273-274, 277, 285-286, 291-292, 301-302, 310-311
Moruzzi, G. 131, 134, 335-336
Motta, S. 338-339
Moustafa, A. 358-359
Mulert, C. 330-331
Müller, J. 42, 45-46, 208-209, 343-344
Munk, H. 72-75, 87, 329, 331-332
Munro, P. 364-365
Musk, E. 376
mutism şi akinezie 24-25, 217-218

N

Naccache, N. 92-93, 331-332

Nader, K. 364
Nagel, T. 42, 93-94, 248-249, 267-268, 308, 332, 370-371
neurochirurgie 18, 32-33, 36-37, 47-48, 51-52, 68-69, 77-78, 102-103, 260-261
neurologie 61-62, 67-68, 74-78
neuromodulare (postsinaptică) versus neurotransmisie (sinaptică) 136-142, 153-154, 217-218, 232-234, 286-287, 355-356, 358-359
inhibare pe termen lung și potențare 139, 232-234, 316-317, 364-365
neuromodulatori/neurotransmițători
acetilcolină 30-32, 36-39, 48-49, 53-55, 140-141, 333-334, 336-337, 341-342, 355-356
aspartat 135, 137
BDNF 365-366
canabinoide 335
CCK 333-334
CRF 333-336
DBI 333-334
dopamină 24-25, 30-31, 35-39, 48- 49, 53-54, 116-117, 123-124, 132- 134, 140-141, 323-326, 335-336, 341-342
estrogen 124-125, 141-142, 333-334
GABA 135, 137
glutamat 135, 137, 333-336
histamină 140-141
LH-RH 333-334
MSH 333-334
neurotensină 333-334
noradrenalină 30-31, 132-134, 140-141, 341-342
NPY 333-334

opioide 123-124, 145-146, 335
orexină 333-334
oxitocină 124-125, 141-142, 333-336
progesteron 125-126, 141-142, 333-334
prolactină 125-126, 141-142, 335-336
serotonină 30-31, 132-134, 140-141, 341-342
substanța P 333-334
testosteron 333-334
vasopresină 333-334
vezi și tipuri de receptori
neuroni piramidali 283-284
neuropsihanaliză 46-38, 54-56
congrese 55-57, 159-160, 286-287, 330-331, 376-377
neuropsihologie
ignoră psihicul 23-25, 47-48, 53-54, 98-99, 167-168
metoda clinico-anatomică 32-33, 131-134, 143-144, 295-297
origini în empirism 70-71, 74-78, 84-85, 223; *vezi și* empirism; Meynert
principiul dublei disocieri 33-34, 322-323
nevoi
conflictuale 117-118, 122-123, 125-126, 199, 290-291
corporale 13-14, 44-45, 57-58, 76- 77, 79-80, 88-89, 96-99, 104-111, 114-117, 142-145, 160- 161, 192, 195-198, 230-231, 352-353, 366-367; *vezi și* funcții autonome/vegetative
emoționale 45, 79-80, 98-99, 106-111, 114-117, 125-126, 142-143, 160-162, 183-185, 197, 199, 205-206, 212-213, 290-291, 372-373
număr mare al 108-109, 184-185, 196, 199-204, 211-212, 219-220, 273, 278, 289-290
prioritizare 108-109, 113-114, 116-118, 145-149, 190, 199-203,

205-206, 211-213, 218-219, 246, 274, 290-293, 361-362, 364-365; *vezi și* context; triunghiul de decizie din mezencefal
și oportunități 108-109, 113-114, 145-149, 290-292, 309-310, 361-362; *vezi și* context; exterocepție
vezi și afecte: versus nevoi; senzații: relația cu nevoile
Newman, J. 92-94, 255-256, 332
Newton, I. 169-170
Nishimoto, S. 284-285, 373-374
noradrenalină *vezi* neuromodulatori/neurotransmițători
Northoff, G. 326-327
Nour, M. 359-360
nuclei și alte structuri cerebrale
accumbens 37-38, 116-117
amigdală 103-104, 116-120, 149-150, 330-331, 333-334, 365-366
aria parafasciculară 335
aria preoptică 335
aria tegmentală ventrală 35-38, 116-117, 132-134, 140-141, 335
cerebel 149-150
coliculi superiori 65-67, 147-149, 204, 249, 287-288, 309-310, 339-340
corpul calos 224-225
corpul geniculat lateral 66-67, 150-151, 224-225, 250, 327-328
formațiune reticulată 337-338
ganglioni bazali 148-150, 192, 270-271, 366-267
gracilis 337
hipotalamus 30-31, 57-59, 115-117, 140-142, 322-323, 337-338, 352-353
locus coeruleus complex 30-31, 132-134

nuclei parabrahiali 295-296, 335-336, 339-340
nuclei pat ai striei terminale 333-336
nucleii prozencefalului bazal 30- 31, 48, 53-54, 140-141, 322-323
nucleii rafeului 30-31, 132-134, 140-141
nucleu cuneat 337
nucleu solitar 104, 337-338, 339-340
nucleu subtalamic 336-337
substanță cenușie periapeductală (PAG) 66-67, 115-116, 118-119, 123-125, 142-148, 203-204, 211-212, 278-279, 287-288, 295-296, 309-310, 333-340, 369-370
substanță neagră 140-141, 336-337
talamus 92-93, 263-264, 322, 324-325, 335
tegument mezopontin 30-31, 33-34, 131, 140-141, 322-323
zona locomotoare din mezencefal 147-148, 339-340; *vezi și* triunghiul de decizie din mezencefal
vezi și cortex; triunghiul de de-cizie din mezencefal; trunchi cerebral
nucleii prozencefalului bazal *vezi* nuclei și alte structuri cerebrale
Nummenmaa, L. 333

O

Oakley, D. 370
Oberauer, K. 365-366
oi 203-205, 208-209, 303-304
Okuda, J. 367
Olds, J. 76
Opatow, B. 327

opioide *vezi* neuromodulatori/neurotransmițători
oportunități *vezi* nevoi
Oppenheim, L. 319
orbire
    corticală 65-67, 72-74, 77-78
    periferică/subcorticală 72-75
orbire a minții *vezi* agnozie
organisme unicelulare 13, 159
orgasm 132-135, 137
oxitocină *vezi* neuromodulatori/neurotransmițători

P

Pace-Schott, E. 324
PAG *vezi* nuclei și alte structuri cerebrale: substanța cenușie periapeductală
Paloyelis, Y. 359
PANICĂ/MÂHNIRE *vezi* emoții de bază
Panksepp, J. 5, 12, 14-15, 38-39, 56-58, 65, 99, 111-113, 123-124, 131, 142-144, 147-149, 151-152, 244-245, 300, 323-324, 326-328, 333-342, 360, 368
Pantelis, E. 319
Parkinson, boală 132
Parr, T. 342, 352
Partridge, M. 323
Parvizi, J. 336, 344
Pasley, B. 284, 374
păsări 111-112, 124-125, 142-143
Pellis, S. 127, 335
Pellis, V. 127, 335

Penfield, W. 77-78, 339
peptide *vezi* neuromodulatori/neurotransmițători
percepție 67, 70, 78, 92-93, 96-97, 148, 183-184, 195, 215-216, 223-224, 253-257
a culorilor 78, 92-93, 107, 224, 227-228, 244-246, 248, 251, 257, 259-260, 265-266, 361, 368
auditivă 62, 78, 138, 152, 175, 208-209, 224, 245-246, 248, 260, 263-264, 278, 339-340, 355, 368, 372
ca halucinație sau fantezie controlată 151, 183-184, 237, 353-354, 360-362
ca realitate virtuală 148-149, 223, 234, 237-238, 245, 339-341, 353-354, 368-369, 374-375
ca testare a ipotezei 150-151, 154, 183-184, 194, 197, 213-214, 228-229, 233-235, 353-354, 360-362
conștientă 151-154, 213, 233-242, 244-246, 248-249, 251, 254-255, 257, 259-260, 265-266, 308, 354, 360-362, 366-369
contextualizează afectul 108-109, 113-114, 118, 145-147, 151-154, 213-214, 220-221, 244-245, 279, 310-311, 368
determinată subiectiv 113-114, 194, 213-214, 248, 251, 259, 265-266, 353-355; *vezi și* context: preconizat; nevoi: și oportunități; predominanță
gustativă 105-106, 113-114, 152, 224, 266-267, 361
inconștientă 62, 65-67, 76, 87, 89-92, 94-95, 148, 183-184, 191, 194, 233-235, 249, 252, 254-257, 270-272, 308, 354, 360-361, 368-369; *vezi și* vedere oarbă
olfactivă 67, 152, 224
somatosenzorială 62, 152, 183-184, 224, 245, 278, 338-342, 368-369
tactilă *vezi* percepție: somatosenzorială
vizuală 62, 66-67, 78, 92-93, 152, 183-184, 203, 208-209, 213-

214, 224-229, 233-235, 244-246, 248-249, 251, 254-255, 257, 259-260, 263-266, 271-272, 278, 284-285, 307-308, 331-332, 339-341, 353- 355, 360, 368, 371; *vezi și* percepție: a culorilor; rivalitate binoculară; suprimare continuă a instantaneelor (*continuous flash suppression*); vedere inversată
vezi și empirism: apercepție; exterocepție; interocepție; psihanaliză: percepție-conștiință
persoana întâi *vezi* subiectivitate: punct de vedere
perspectivă *vezi* subiectivitate: punct de vedere
PET *vezi* tomografie cu emisie de pozitroni
Petkova, V. 368
Pezzulo, G. 353, 362
Pfaff, D. 313-316, 337
Picard, F. 352
Planck, M. 345
plasticitate 140, 193, 236, 355
vezi și consolidare/reconsolidare
plăcere/neplăcere *vezi* sentimente: valența
Popper, K. 29, 321-322
Porac, C. 366
Porte, J.-P. de la 41-42
Posner, M. 327
Pouget, A. 352
precizie (încredere, varianță inversă, amplificare postsinaptică, forță a semnalului) 138, 140, 150, 152-154, 206-221, 229-231, 234, 236, 242, 246, 273, 290-292, 300-301, 310-311, 326, 355-358, 363-364, 366
ca activare/selecție 138, 140, 151-152, 206-207, 211-212, 218, 233-234, 237, 242, 358-359

vezi și demonul lui Maxwell; neuromodulare versus neurotransmisie
predictibilitate *vezi* incertitudine; probabilitate
predicție
erori *vezi* semnale de eroare
în experimentul științific 20-21, 27-29, 37-38, 80, 99, 154, 259-260, 293-295, 301-302; *vezi și* știință: reguli
în inferența activă 149-154, 159, 175-176, 180-181, 194-195, 206-217, 282-283, 296, 361
în inferența perceptivă 183-184, 193-195, 197, 227-229, 233-235, 282-283, 354, 360-362, 366-369
înnăscută 116-117, 149, 191, 196-197, 199-200, 234-235, 310, 354, 375; *vezi și* instinct și reflex
suprimă semnalele senzoriale 150-151, 183-184, 193, 233-235, 361-362
vezi și Bayes: regula, teorema; incertitudine; memorie: ca predicție; probabilitate
predominanță 144-145, 147-149, 154-155, 197, 204-206, 211-213, 237-238, 246, 273, 337-338, 341-342, 367-368
definiție 213-214
vezi și nevoi: prioritizare; percepție: determinată subiectiv
Pribram, K. 325
primate 112, 118, 142, 243, 282-284
prioritizare *vezi* nevoi: prioritizare
probabilitate 164-165, 179, 181-184, 245, 315, 344, 351
posterioară/anterioară 183-184, 195, 206-208, 366; *vezi și* Bayes: regula
problema altor minți *vezi* filosofie
problema dificilă 12-14, 247-275, 286-288, 293-294, 299-302, 308
problema fixării *vezi* filosofie

problema minte-corp *vezi* filosofie
„Proiect de psihologie” *vezi* Freud
prosopagnozie 227-228
prozencefal 29-32, 38-39, 44-45, 62-63, 88-89, 141-150, 152-154, 173-174, 197-198, 205-206, 209-210, 238-239, 309-310, 329-330, 336-340, 365-368
vezi și cortex
pseudocomă 68-69
psihanaliză 12-13, 21, 28-32, 41, 43- 45, 47-48, 53-56, 102, 106-107, 113, 123, 277
apărare 43, 356-357
cenzură 43
conținut manifest/latent 28-29, 46
dorință 28-29, 31, 37, 47, 54, 58, 120, 257, 274, 326
ego 58-59, 327, 343
energie psihică 45-46, 57, 159-160, 363
inconștient 15, 43, 58-59, 72-73, 88-90, 95-96, 107-108, 270, 277, 331, 341
libido 115, 326; *vezi și* pulsiune
metapsihologie 44-46, 57-58, 324, 341
metoda asocierii libere 28-29
percepție-conștiință (P-Cs) 15, 44-45, 57-59, 64, 87-90, 95-96, 328, 331-332, 341, 357, 363
preconștient (Pcs) 88-90, 331, 363
principiul plăcerii 57, 97-98, 357
procese primare 343, 356
procese secundare 152, 343, 348, 356, 363; *vezi și* amânare
pulsiune *vezi* pulsiune
represie 43, 356, 364
rezistență 43, 363-364

se 57-59, 341
terapie prin vorbire 243, 293-294
testarea realității 54-56, 150-151, 326, 356
vezi și Freud
psihiatrie/psihofarmacologie 29, 31- 32, 36, 54-55, 75, 83-85, 98-99, 133, 159, 218-219, 326, 330
psihoză 25, 36, 49, 132-133, 150
pulsiune 24-25, 28-29, 31-32, 37-39, 44-46, 54-55, 57-58, 76, 88-90, 106-108, 116-117, 128, 141-143, 147-148, 158-159, 184-185, 196-197, 213-214, 229-231, 271-272, 313-314, 325-326, 329-330, 337-338, 363-364
vezi și travaliu

Q

Qin, P. 331
qualia *vezi* filosofie: qualia
Queen Square, Londra 158

R

Ramachandran, V.S. 326-327, 359-360
Rance, D. 372-373
Ratliff, F. 365-366
recompensă/pedeapsă *vezi* behaviorism
reconsolidare *vezi* consolidare/reconsolidare
Reggia, J. 374-376
REM *vezi* somn
Remmelzwaal, L. 373
reprezentare 11, 91-92, 94-97, 99, 148-150, 174-175, 200, 217-218,

223, 322-323, 348
origine 174-175
vezi și filosofie: pe cine vizează intenția; Meynert: imagini mnezice
retină 193-194, 224-225, 227-228, 250-251, 257-260, 265-266, 328-329, 359-360, 371-372
Reymond, E. du Bois 42, 307, 324-325
Richardson, R. 324-325
Riggs, L. 223-235, 365-366
rivalitate binoculară 227-229, 361-362
vezi și suprimare continuă a instantaneelor (*continuous flash suppression*)
Roepstorff, A. 242-243, 367-368
Rolls, E. 37-38, 201-202, 323-324, 354-355
Rosch, E. 355-356
Rosenthal, D. 330-331
Rovelli, C. 345, 347
Rowley, M. 326-327
Royal London, spital, Whitechapel 48
Runes, D. 328-329

S

Sacks, O. 24-26, 37-38, 53-54, 56, 95-96, 98-99, 167-168, 217-218, 321, 326-327
Sacktor, T. 326-327
Saling, M. 102-103, 325-326, 330-331
saturație *vezi și* homeostază; sentimente 98-99, 106-109, 114-115, 230-232
scala de comă Glasgow 68-69, 80-81, 134-135, 327-328

Schacter, D. 239-241, 326-327, 367-368
Schafe, G. 364-365
Schindler, R. 224-225
Schore, A. 326-327
Schultz, W. 364-365
Schwartenbeck, P. 35
Searle, J. 250, 279-280, 282-283, 369-373
selecție naturală 109-110, 114-115, 168-170, 290-291, 316-317
Semenza, C. 326-327
semnale de eroare 106-111, 141-143, 145-148, 150-151, 153-155, 169-171, 175-176, 183-185, 188-191, 193-198, 201, 204-212, 214-220, 234-235, 237-238, 246, 282-283, 309-311, 326-327, 351-352, 360-362, 366-367
vezi și surprindere
sentimente
ca semnale de eroare 106-111, 154-155, 160-162, 230-232, 308-309, 356-358; *vezi și* homeostază; saturație; semnale de eroare
definiția termenului 95-96, 321-322
forța cauzală 13-14, 96-99, 105-110, 197, 201, 212-214, 254-256, 266-268, 270-273, 309-310; *vezi și* alegere
ghidează rezolvarea problemelor 109-110, 112-113, 116-118, 120-121, 125-126, 128, 151-155, 200, 205-214, 229-232, 238-239, 246, 309-310
importanță 12, 14-15, 35-36, 52- 53, 80-81, 96-99, 105-110, 118-119, 124-125, 128-129, 321- 322, 366-367
negate de behaviorism 19-20, 97-99, 201, 219-220, 367-368
nu sunt o funcție a cortexului 58-59, 64-67, 69-70, 79-81; *vezi și* eroarea privitoare la cortex
relația cu nevoile 44, 57-59, 76-77, 78-80, 88-89, 98-99, 104-

110, 113-115, 132-134, 144-149, 184-185, 197, 200-201, 211-214, 230-232, 236-237, 271-273, 309-310, 362, 366-367; *vezi și* homeostază; nevoi: prioritizare
sunt cu necesitate conștiente 57-58, 94-96, 102, 107-109, 271-274, 308-309, 321-322
sunt forma fundamentală a conștiinței 132-134, 151-152, 154-155, 197-198, 203-204, 217-218, 220-221, 246, 271-272, 285-287, 308-309
sunt o funcție a trunchiului cerebral 132-134, 144-146
valența 13, 57, 98-99, 104-105, 108-111, 119-120, 146-147, 157, 200-201, 203-204, 219-220, 244-245, 291-293, 296-297, 309-310, 332-333, 364-365;
versus „emoție inconștientă” 95-96, 102, 121, 271-272, 330-332
versus inteligență 13-14, 220-221, 285-286, 288-289, 293-295, 298-299, 304-305, 330-331
versus la ce se referă 90-92, 94-96, 103-106, 121-122, 128-129, 235-236, 304-305
vezi și afecte; conștiință
serotonină *vezi* neuromodulatori/neurotransmițători
Sessions, R. 354-355
Seth, A. 358-359
Shallice, T. 324-327
Shannon, C. 166-169, 265, 315-316, 345-346
Shapiro, T. 326-327
Sharf, B. 323-324
Sherman, S. 358-359
Shewmon, D. 327-328
Shiffrin, R. 370-371
Shipp, S. 351-352

Shock, J. 319, 372-373
Simons, D. 355-356
simțuri *vezi* percepție
sincopă vasovagală 103-105, 146-147
vezi și funcții autonome/vegetative: reflex vasovagal
sine 23-25, 42, 52-53, 56-57, 71-72, 78-80, 84-85, 368-369
procesare legată de sine 78, 83-84
SINELE primar *vezi* triunghiul de decizie din mezencefal
vezi și psihanaliză: ego; subiectivitate
Singer, W. 326-327, 358-359
sistem endocrin 140-142, 195-196
sisteme cu autoorganizare 157-159, 169-182, 190-191, 200-202, 205-206, 219-220, 230-231, 237-238, 267-268, 273-274, 288-290, 291-292, 309-311, 343, 347
care se autovalidează 180-182, 214-216, 219-221, 269-270, 288-289, 293, 311-312
definiția sintagmei „care se autovalidează” 351-352
vezi și homeostază
sistemul reticulat activator 39, 131-138, 140-145, 151-155, 162, 173-174, 209-212, 218-219, 232-233, 278-279, 308-309, 322-323, 336-338, 364-366
vezi și activare; circuite; nuclei și alte structuri cerebrale
situații neprevăzute *vezi* incertitudine
Skinner, B.F. 97-98, 332-333
Skotnes, P. 319
Societatea de Fizică din Berlin 42, 45-46, 178-179, 311-312, 343-344
Solms, K. Kaplan 326-327, 330
Solms, L. 17-19, 101-102, 104
Solms, M. 322-336, 341-342, 344, 354-355, 357-362, 364-365, 367-

369, 374-376
somn 26-31, 48, 65, 84-85, 99, 116-117, 134-140, 166-167, 199, 214-215, 238-239, 290-291, 301-302, 323-324, 336-337, 366-367
REM/non-REM 27-35, 37-38, 48, 95-96, 128, 134-135, 137-138, 322-324, 333-334, 336-337
vezi și funcții autonome/vegetative: ciclul somn – stare de veghe
Sperry, R. 90-92, 94-96
spike trains 139, 180-181, 224-225, 245
Squire, L. 331-332
stare de conștiență minimă 166-167
stare de repaus *vezi* mod implicit
stare de veghe 26-27, 62, 67-68, 134, 138-140, 142-144, 314-315
vezi și activare; conștiință: experiență versus starea de veghe
stare de veghe vegetativă *vezi* stare vegetativă
stare vegetativă 61-65, 67-70, 75-76, 134-135, 142-143, 166-167, 285-286
Stein, T. 362
Steiner, R. 326-327
Stephan, K. 352-353, 359-360
Sterzer, P. 362
stimul/răspuns 21-24
stimulare cerebrală profundă 111-112, 132-134, 144-146, 295-297, 330-331, 336-339
stimulare optogenetică 259-260, 372
Stoerig, P. 340-341
Strachey, J. 45, 325-327
Strawson, G. 331-332
subiectivitate 11, 13, 19-21, 24-25, 33-34, 42, 44-55, 95-97, 109-

110, 174-175, 182, 201, 204, 218-219, 248, 251-252, 265-266, 279, 291-293, 296-297, 308, 310
ipseitatea 174, 176, 219-220, 292-294, 309-310, 372-373
origine 174, 218-220, 252-253, 258-260, 263-266, 274-275, 278-279, 308-309
punct de vedere 11-12, 49-53, 148-149, 174, 176, 218-220, 237-240, 256-257, 259-260, 265-266, 288-289, 291-292, 310, 339-341, 348-349, 368-371
putere cauzală; *vezi și* liber-arbitru 25-26, 110-111, 212-214, 250-252, 266-268, 270-272, 309-310
rapoarte introspective; *vezi și* limbaj 20-22, 24-29, 31-32, 34-35, 41-42, 77-78, 84-85, 95-97, 201, 225-227, 256-257, 293-294, 328-331, 333-334, 396-370,
vezi și adresare și răspuns la întrebări; filosofie; sine
substanța cenușie periapeductală *vezi* nuclei și alte structuri cerebrale
Sulloway, F. 325-326
supa primordială *vezi* Friston
supraajustare *vezi* model generativ: complexitate
suprimare continuă a instantaneelor (*continuous flash suppression*) 240-242, 367-368; *vezi și* rivalitate binoculară
surprindere 179-181, 230-231, 316-317, 349-352
surpriză 110-111, 151-152, 154, 244, 367-368
vezi și afecte: senzoriale
Szpunar, K. 367-368

Ș

știință, reguli 19-20, 29-30, 41, 68-69, 99, 258, 278, 292-293, 296-297, 299-302, 321-322, 372-373

vezi și empirism; predicție: în știință; testarea ipotezei: în știință

T

tegment mezopontin *vezi* nuclei și alte structuri cerebrale
teoria informației integrate *vezi* conștiință: teoriile
teoria reintrării *vezi* conștiință: teorii
teoria spațiului de lucru global *vezi* conștiință: teorii
terapie electroconvulsivă (ECT) 231-232, 364-365
termodinamică 159-160, 162-166, 178-180, 236-237, 274, 315-316, 345-346, 349
vezi și lege: a termodinamicii
termoreglare *vezi* funcții autonome/vegetative
testarea ipotezelor
în percepție 150-152, 154-155, 183-184, 194, 213-214, 228-229, 233-235, 277-278, 360-362
în știință 19-20, 26-31, 36-37, 54- 55, 64-65, 67-58, 80-81, 259-260, 277-278, 286-288, 293, 299-302; *vezi și* știință, reguli
vezi și Bayes: regula
testul Turing 293-295, 375-376
Thomas, K. 365-366
Thompson, E. 355-356
Thompson-Schill, S. 367-368
Thorndike, E. 97-99, 332-333
Tinbergen, N. 313
tipuri de receptori 138, 140-142
α-MSH 333-334
μ-opioid 335-336
nicotinic 31

vezi și neurotransmițători/neuromodulatori
tomografie cu emisie de pozitroni (PET) 23, 47, 133-134, 137-138
Tononi, G. 93-94, 331-332, 359
Tossani, E. 334-336
Tozzi, A. 346, 352-353
Tranel, D. 326-327, 330-331, 333-334
travaliu 45-46, 57-58, 88-89, 105-108, 116-117, 141-142, 146, 150, 152-153, 160-165, 181-182, 184-185, 197, 229-232, 235-239, 246, 286-287, 289-290, 297-298, 361-363
triunghiul de decizie din mezencefal 146-151, 153-154, 162, 173-174, 196, 203-206, 209-210, 218-219, 309-310, 338-339, 364-365
trunchi cerebral 14-15, 29-31, 34-36, 38-39, 56-59, 65-67, 69-70, 131-137, 141-146, 193-194, 244-245, 249-250, 309, 322-323, 331-332, 336-337, 339-340, 352-353, 359-360, 261-262
bulb rahidian 337-340
canal central 142-143, 145-146, 352-354
definiția termenului 322-323
diencefal 142-143, 193-194, 338-340; *vezi și* nuclei și alte structuri cerebrale: hipotalamus; talamus
mezencefal 132-134, 142-148, 278-279, 294-295, 303-304, 338-340; *vezi și* triunghiul de decizie din mezencefal
punte 29-31
vezi și activare; nuclei și alte structuri cerebrale; sentimente; sistemul reticulat activator
Tsakiris, M. 256-257
Tse, P. 365-366
Tsodyks, M. 370-371
Turing, A. 375-376
Turnbull, O. 293-294, 326-327, 333-334, 374-376
Turner, M. 326-327

U

Uhlhaas, P. 359
University of the Witwatersrand, Johannesburg 32-33, 41-42
Urbach-Wiethe, boală 119-121

V

valență *vezi* sentimente: valența
valori 13-14, 104-106, 108-109, 124, 128, 145-147, 151-152, 164-165, 200-201, 210-211, 219-221, 288-292, 296-299, 303-304
  succes reproductiv 113-114, 296
  vezi și selecție naturală
Varela, F. 213-214, 348-349, 355-356
varianță *vezi* precizie
vedere inversată 224-228
vedere oarbă 65-67, 77-78, 148-149, 327-328, 333-334, 340-341
Velmans, M. 326-327
Venkatraman, A. 337-338
vertebrate 111-112, 142-143, 151-152, 200, 219-221, 278, 295-296, 303-304
Vertes, R. 341-342
viață *vezi* fizică; homeostază; mortalitate; selecție naturală
vise 12, 26-41, 46-49, 53-54, 73-75, 83-84, 95-96, 116-117, 120-121, 128, 133-137, 217-218, 237-239, 284-285, 322-326, 229-230, 359-360, 366-367, 374
volițiune *vezi* intenţionalitate
voluntar versus involuntar 21-22, 61, 75-76, 87, 107-110, 113-115, 117-118, 120-122, 128, 142-143, 147, 149-151, 181-182, 191-192, 200, 206, 229-230, 237-238, 277, 294-296, 310, 314-315, 363

definiția termenului „voluntar” 108-109
vezi și Meynert

W

Walker, M. 336-337
Wang, X. 359-360
Ward, E. 240-241, 367-368
Watt, J. 375-376
Watts, J. 35-36
Weiskrantz, L. 66-67, 327-328
Weizenbaum, J. 375-376
Wernicke, C. 24-25, 74-75
Westhuizen, D. van der 334-336, 362, 373
Wheeler, J. 167-168, 176, 220-221, 245, 265-266, 346-347, 371-372
White, B. 148-149, 340-341
Whitty, C. 83-84, 330-331
Widlöcher, D. 326-327
Wiener, N. 169-170
Wilbrand, H. 72-74, 328-330
Wolpert, D. 359
Wyzinski, P. 322-323

Y

Yang, E. 362
Yovell, Y. 333-334
Yu, C. Kai-Ching 323-324

Z

Zahavi, D. 371-372
Zald, D. 362
Zare, M. 346, 35-353
Zeki, S. 224, 321-322, 360
Zellner, M. 325-327
Zeman, A. 67, 134, 327-328
Zenon (din Elea) 345
zgomot *vezi* precizie
Zhou, T. 334-336, 367-368
zombi *vezi* filosofie

* Numărul paginilor se referă la ediţia tipărită a cărţii şi la ediţia digitală în format PDF.